

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL
DEPARTAMENTO DE INFORMÁTICA



Previsão de consumos de eletricidade e gás em Portugal – Análise, Adaptação e Evolução

Carolina Linhares de Lima de Carvalho Seabra

Mestrado em Gestão de Informação
Especialização em Gestão e Análise de Dados

Trabalho de projeto orientado por:
Professor Doutor Fernando José Araújo Correia da Ponte Sequeira
Professor Doutor Pedro Miguel Frazão Fernandes Ferreira

Agradecimentos

Quero começar por agradecer aos meus orientadores, Professor Fernando Sequeira e Professor Pedro Ferreira, pelo apoio dado ao longo do projeto e por ajudarem-me a ultrapassar todas as dúvidas que foram aparecendo.

Também quero agradecer ao Rui Coelho, ao Eng. Milheiro Batista e a toda a equipa do departamento ACER – Estudos e Estatísticas da REN, pela oportunidade deste projeto e por me aceitarem. E principalmente pelo apoio e ajuda a ultrapassar todos os obstáculos.

À minha família, por serem o meu suporte. Especialmente à minha mãe, com as suas grandes palavras, nunca me deixou desmotivar.

Finalmente, aos meus grandes amigos, Patricia Gi Fernandes, Joana Fernandes e Nuno Pinheiro, por me darem força para continuar e serem uns grandes apoiantes. E a todos os meus colegas e amigos que estiveram presentes para mim, como a Ana Carolina Guerreiro e Rui Nogaró.

Resumo

A empresa REN - Redes Energéticas Nacionais é responsável pelo transporte de eletricidade e gás natural e ainda pela gestão técnica dos Sistemas Nacionais de Eletricidade e dos Sistemas Nacionais de Gás Natural. A empresa também assegura o balanço entre os valores de produção e consumo de eletricidade e por essa razão detém um modelo de previsão para tal. Um modelo de previsão simples do tipo auto-regressivo ou do tipo alisamento exponencial não é por si só suficientemente preciso, pelo que exige procedimentos envolventes, como por exemplo correções aos dados. Este processo, que consiste num modelo de Holt-Winters, inclui algumas variáveis exógenas necessárias, que influenciam o consumo de eletricidade no dia-a-dia da sociedade, como por exemplo a temperatura e o tipo de dia da semana e do ano.

Este projeto baseia-se essencialmente no tratamento dessas variáveis e no estudo de inclusão de outras que influenciam o consumo de eletricidade da população portuguesa. Começando com uma atualização da variável temperatura, pois o clima tem vindo a alterar-se ao longo dos anos, e com um ajuste na definição do processo envolvente desta variável, pretende-se que no fim continuem a existir erros de previsão baixos. Com o modelo atualizado para os novos valores, tem-se como objetivo estudar a inclusão de uma nova variável exógena no modelo utilizado como base deste projeto. Deste modo, tenciona-se analisar a inclusão da variável luminosidade, fazendo um estudo de forma a verificar se esta influencia o consumo, ou seja, se existe uma relação entre o consumo e tal medida. Por fim, intenta-se adaptar o modelo, apenas com a atualização da variável temperatura, ao consumo de gás.

A apresentação dos resultados consiste essencialmente na medida de erro RMSE (Root Mean Squared Error) da previsão. Também é feita uma comparação ao erro percentual anual, erro utilizado e estudado pela empresa. Estes erros são comparados no ajuste do processo da variável temperatura e na adaptação do modelo ao consumo de gás. No caso do estudo da inclusão da variável luminosidade são apresentados os resultados da correlação entre esta e o consumo de eletricidade.

Palavras-chave: previsão, consumo, eletricidade, gás.

Abstract

The company REN – Redes Energéticas Nacionais is responsible for electricity and natural gas transportation and for the overall technical management of the Portuguese National Electricity and the National Natural Gas Systems. REN must also ensure the balance between production and consumption of electricity, and, for that reason it requires a forecasting model. However, that prediction is not resourceful enough in the case of simple autoregressive or exponential smoothing models, therefore a more complex process and model is required to generate this prediction. This process, which consists of a Holt-Winters model, includes some necessary exogenous variables that have a strong influence in electricity consumption, like the weather temperature and day specifications, like weekday, holyday or seasonality.

This project is based on the processing of those variables and in the study of others that may influence the electricity consumption. Having that in mind, this study will start with an update of the variable temperature, because the weather has been changing along the years, in order to maintain low error in the forecast. With new values for this variable, another goal is to study the inclusion of a new variable, luminosity. It is believed that this measure influences the electricity consumption and because of that we are going to study the relation between these two, in order to verify its influence. In the end, we will also analyze the time-series of gas consumption, with the aim of adapting the model to it.

The presentation of results consists in comparing the RMSE (Root Mean Squared Error) of the predictions. Also, we will provide a comparison between the annual percentage errors, which are compared in the update of the temperature variable and in the adaption of the model to the gas consumption. In the luminosity case study a correlation analysis to the consumption is presented.

Keywords: prediction, consumption, electricity, gas.

Índice

Agradecimentos	ii
Resumo.....	iii
Abstract.....	iv
Índice.....	v
Índice de figuras	vi
Índice de tabelas	viii
Capítulo 1 – Introdução.....	1
I. Apresentação.....	1
II. Objetivos.....	3
III. Estrutura do documento	4
Capítulo 2 – Método de Previsão	6
I. Séries Temporais	6
II. Modelos de Previsão	6
a. Modelo de Alisamento Exponencial	7
b. Modelo de Holt	8
c. Modelo de Holt-Winters	8
d. Tipos de Tendência	9
III. Processo de previsão da empresa.....	10
a. Série de consumo de eletricidade.....	10
b. Correção Temperatura.....	12
c. Modelo.....	15
d. Efeito Calendário	16
e. Efeito da Temperatura.....	17
Capítulo 3 – Metodologia	18
I. Coeficiente de correlação e dependência linear	18
II. Regressão linear múltipla	18
Capítulo 4 – Resultados.....	22
I. Definição de metodologia de reparametrização	22
a. Temperatura Normal	22
b. Gradientes	29
II. Luminosidade	34
III. Gás	41
a. Relação.....	41
b. Adaptação do modelo	46
Capítulo 5 - Conclusão.....	52
Referências.....	54
Anexos.....	55

Índice de figuras

Figura 1.1 – Consumo de eletricidade em Portugal do ano 2015.	3
Figura 1.2 – Padrão semanal de consumo de eletricidade em Portugal.	3
Figura 2.1 – Exemplo sazonalidade multiplicativa.	9
Figura 2.2 – Série de consumo de eletricidade nos anos 2006 até 2015.	11
Figura 2.3 – Série de consumo de eletricidade do ano 2015.	11
Figura 2.4 – Periodograma da série referente ao consumo de eletricidade em Portugal nos anos 2006 até 2015.	11
Figura 2.5 – Periodograma da série referente ao consumo de eletricidade em Portugal no ano de 2015.	12
Figura 2.6 – Relação entre os desvios de consumo e desvios de temperatura na semana 15.	13
Figura 2.7 – Exemplo três gradientes.	14
Figura 2.8 – Exemplo caso $\phi_{ikj} < -2$	14
Figura 2.9 – Exemplo caso $-2 \leq \phi_{ikj} < 2$	15
Figura 2.10 - Exemplo caso $\phi_{ikj} \geq 2$	15
Figura 3.1 – Diagrama de dispersão entre a temperatura e o consumo de eletricidade nos anos de 2006 até 2015.	18
Figura 4.1 – Temperaturas normais com escala de 18 de julho a 17 de julho do ano a seguir.	22
Figura 4.2 – Temperaturas normais com escala de 1 de janeiro a 31 de dezembro.	23
Figura 4.3 – Temperaturas médias com escala de 1 de janeiro a 31 de dezembro.	23
Figura 4.4 – Ajustamentos polinomiais de grau 8, 10, 12 e 13.	25
Figura 4.5 – a) Pontas iniciais das curvas dos ajustamentos polinomiais de grau 8, 10, 12 e 13.	25
b) Pontas finais das curvas dos ajustamentos polinomiais de grau 8, 10, 12 e 13.	25
Figura 4.6 – a) Pontas iniciais das curvas dos ajustamentos polinomiais de grau 8, 10 e 12.	26
b) Pontas finais das curvas dos ajustamentos polinomiais de grau 8, 10 e 12.	26
Figura 4.7 – Ajustamentos polinomiais de grau 6 e 8 para o conjunto de dados dos últimos 10 anos.	27
Figura 4.8 – Ajustamentos polinomiais de grau 8 e 10 para o conjunto de dados dos últimos 20 anos.	27
Figura 4.9 – Ajustamentos polinomiais do conjunto de todos os dados, conjunto últimos 10 anos e conjunto últimos 20 anos com os graus mais adequados respetivos.	27
Figura 4.10 – a) Pontas iniciais dos ajustamentos polinomiais da figura 4.9.	28
b) Pontas finais dos ajustamentos polinomiais da figura 4.9.	28
Figura 4.11 – Ajustamentos polinomial de grau 10 para o conjunto dos últimos 20 anos.	28
Figura 4.12 – Ajustamento à junção das pontas do anterior ajustamento polinomial.	29
Figura 4.13 – Temperaturas normais finais.	29
Figura 4.14 – Evolução e convergência do gradiente 1.	31
Figura 4.15 – Evolução e convergência do gradiente 2.	31
Figura 4.16 – Evolução e convergência do gradiente 3.	32
Figura 4.17 – Efeito fator de correção da semana 1.	33
Figura 4.18 – Efeito fator de correção da semana 52.	33
Figura 4.19 – Série de consumo de eletricidade de 2009 até 2015.	34
Figura 4.20 – Periodograma da série de consumo de eletricidade apresentado na figura 4.19.	35
Figura 4.21 – Série de luminosidade registada de 2009 até 2015.	35
Figura 4.22 – Série de percentagem de luminosidade registada de 2009 até 2015.	35
Figura 4.23 – Série de percentagem de luminosidade registada no ano 2015.	36

Figura 4.24 – Diagrama de dispersão entre consumo de eletricidade e percentagem de luminosidade em dias úteis.	36
Figura 4.25 – Diagrama de dispersão entre o consumo de eletricidade e percentagem de luminosidade de todas as centrais em dias úteis.	38
Figura 4.26 – Diagrama de dispersão entre o consumo de eletricidade e percentagem de luminosidade da central em Lisboa.	38
Figura 4.27 – Percentagem de luminosidade média diária anual.	39
Figura 4.28 – Curvas dos ajustamentos polinomiais de grau 4, 5, 6 e 7.	40
Figura 4.29 – a) Pontas iniciais dos ajustamentos polinomiais de grau 6 e 7.	40
b) Pontas finais dos ajustamentos polinomiais de grau 6 e 7.	40
Figura 4.30 – Ajustamento polinomial de grau 7 e médias diárias de percentagem de luminosidade.	41
Figura 4.31 – Série de consumo de gás natural de 2013 até 2015.	41
Figura 4.32 – Periodograma da série de consumo de gás.	42
Figura 4.33 – Periodograma da série de temperatura registada.	42
Figura 4.34 – Diagrama de dispersão entre consumo de gás e temperatura.	43
Figura 4.35 – Diagrama de dispersão entre o consumo de gás e as temperaturas, dividido pelas estações do ano.	44
Figura 4.36 – Diagrama de dispersão entre o consumo de gás e as temperaturas, dividido pelo tipo de dia.	44
Figura 4.37 – Consumo médio por tipo de dia.	44
Figura 4.38 – a) Diagrama de dispersão entre o consumo de gás e temperatura para os dias úteis.	45
b) Diagrama de dispersão entre o consumo de gás e temperatura para o inverno.	45
Figura 4.39 – a) Diagrama de dispersão entre o consumo de gás e temperatura para os dias úteis dividida pelas estações do ano. b) Diagrama de dispersão entre o consumo de gás e temperatura para o inverno dividida pelo tipo de dia.	46
Figura 4.40 – Evolução e convergência do gradiente 1.	48
Figura 4.41 – Evolução e convergência do gradiente 2.	48
Figura 4.42 – Evolução e convergência do gradiente 3.	48
Figura 4.43 – Efeito fator de correção da semana 1.	49
Figura 4.44 – Evolução e convergência do gradiente 1.	50
Figura 4.45 – Evolução e convergência do gradiente 2.	50
Figura 4.46 – Evolução e convergência do gradiente 3.	51
Figura 4.47 – Efeito fator de correção da semana 52.	51

Índice de tabelas

Tabela 4.1 – Significância dada pelo teste t-student aos coeficientes para cada ajustamento polinomial de grau 5, 10, 15 e 20.	24
Tabela 4.2 – Coeficiente de determinação (e ajustados) para cada ajustamento polinomial de grau 5, 10, 15 e 20.....	24
Tabela 4.3 - Coeficiente de determinação (e ajustados) para cada ajustamento polinomial de grau 8, 10, 12 e 13.....	25
Tabela 4.4 – Significância dada pelo teste t-student aos coeficientes para o ajustamento polinomial de grau 12, para os diferentes conjuntos de dados dos últimos 10 anos e dos últimos 20 anos.	26
Tabela 4.5 – RMSE de previsão e graus obtidos para o ajustamento polinomial dos três gradientes, para cada intervalo testado.....	30
Tabela 4.6 – RMSE de previsão para cada combinação de temperaturas normais e gradientes.	31
Tabela 4.7 – Erro anual em percentagem de previsão de 2003 até 2015.....	33
Tabela 4.8 – RMSE de previsão e erro anual em percentagem de previsão sem valores históricos de 2015.	33
Tabela 4.9 – Coeficiente de correlação entre consumo de eletricidade e percentagem de luminosidade, dividido por tipo de dia.....	37
Tabela 4.10 – Coeficiente de correlação para cada análise, dividida por tipo de dia.	37
Tabela 4.11 - Significância dada pelo teste t-student aos coeficientes para cada ajustamento polinomial de grau 10 e 15.	39
Tabela 4.12 – Coeficiente de determinação (e ajustados) para cada ajustamento polinomial de grau 10 e 15.....	39
Tabela 4.13 – Coeficiente de correlação total e para cada ano.	43
Tabela 4.14 – Coeficiente de correlação para cada ano diferenciado o conjunto de temperaturas normais.	43
Tabela 4.15 – Coeficiente de correlação por tipo de dia.	45
Tabela 4.16 – Coeficiente de correlação por estação do ano.	45
Tabela 4.17 – RMSE de previsão para cada teste, diferenciado o conjunto de temperaturas normais. .	47
Tabela 4.18 – Erro anual em percentagem de previsão de 2013 até 2015.....	47
Tabela 4.19 – Valores de correção direta ao consumo para a semana 1 no caso de um desvio de temperatura maior que 2.	49
Tabela 4.20 – RMSE de previsão para cada teste diferenciado o conjunto de temperaturas normais. ..	49
Tabela 4.21 – Erro anual em percentagem de previsão de 2013 até 2015.....	49
Tabela 4.22 – Valores de correção direta ao consumo para a semana 52 no caso de um desvio de temperatura menor que -2.....	51

Capítulo 1 – Introdução

I. Apresentação

Hoje em dia, a energia é considerada um bem essencial para a sobrevivência do ser humano. No contexto deste projeto, as fontes de energia dividem-se entre as energias não renováveis e as energias renováveis, as quais podem ser transformadas em eletricidade e gás natural.

As energias não renováveis são fontes de energia que dependem de processos em escala de tempo geológico que incluem petróleo, gás natural, carvão mineral e energia nuclear. Isto significa que estas dependem da existência de combustíveis fósseis e das suas reservas, que uma vez esgotadas, não podem ser regeneradas, pois as quantidades são limitadas. Necessariamente são transformadas numa energia secundária, como a eletricidade ou a gasolina. Têm como vantagem principal a apresentação de um elevado rendimento energético (daí serem a maior fonte para estas energias secundárias) mas também têm as suas desvantagens: são consideradas energias sujas, com muitas consequências negativas para o ambiente e para a sociedade e têm sido bastante prejudiciais ao ponto de existirem países a preferir as energias renováveis às não renováveis, pois apesar do seu custo elevado, estas acabam por trazer maiores benefícios a longo prazo.

As energias renováveis são fontes de energia que vêm de recursos naturais que, por sua vez, são naturalmente reabastecidos, isto é, a sua quantidade é ilimitada pelo que não se esgotam, ao contrário das energias não renováveis. Dentro destas energias, tem-se a energia solar, que provém do sol, a energia eólica, que subsiste do vento, a energia hidráulica, com origem na água, sendo estas as que apresentam uma maior indústria. A energia solar é a energia obtida pela luz do sol, através da tecnologia de energia fotovoltaica que, por meio de painéis fotovoltaicos converte a energia da luz diretamente em energia elétrica.

Com várias formas de energia, permitiu-se gerar eletricidade a um nível muito elevado. A eletricidade, hoje em dia, é utilizada pelas grandes indústrias e empresas, como por exemplo, para o funcionamento das máquinas e computadores. Em seguimento, esta também é utilizada por grande parte da humanidade, no seu quotidiano, principalmente nos utensílios tecnológicos, como telefones e televisões, e ainda nos eletrodomésticos e iluminação. Deste modo conseguimos compreender que a eletricidade está sempre presente na vida das pessoas nos tempos modernos, o que contribuiu para o seu desenvolvimento e lhe deu a capacidade de se tornar um dos maiores negócios a nível nacional.

Juntamente com a eletricidade, o gás natural também construiu um negócio poderoso, sendo uma energia não renovável, cuja maior utilização se centra no seu uso como combustível, seja na indústria, em casas ou em automóveis. Ao contrário dos restantes combustíveis fósseis, este é o que apresenta mais vantagens em relação ao ambiente, pois é uma das fontes de energia mais limpas e seguras, em comparação com as restantes.

A eletricidade em Portugal surgiu durante o século XIX, onde várias empresas foram criadas para melhorar a infraestrutura do país. Ao longo do tempo, e com o aumento do consumo, sendo a maior fonte de energia as energias não renováveis, o país conseguiu atingir novas metas ecológicas - atualmente mais de 50% da eletricidade consumida no país é obtida através da energia eólica. O gás natural só foi introduzido mais tarde, no fim do século XX, de forma a reestruturar o setor energético do país.

Assim, a energia, ao ser considerada um bem precioso, permite a existência e a gestão de um mercado bastante desenvolvido e crescente que envolve várias empresas, em diferentes campos como a produção, distribuição ou comercialização. Uma dessas empresas é a REN - Redes Energéticas Nacionais, que tem como principais áreas de negócio, o transporte de eletricidade e gás natural por todo o território Português.

Em detalhe, as duas grandes funções da REN consistem no transporte de eletricidade em muito alta tensão e a gestão técnica global do Sistema Elétrico Nacional, e ainda no transporte de gás natural em alta pressão e a gestão técnica global do Sistema Nacional de Gás Natural. Além destas duas, a REN também está presente nas energias renováveis e nas redes de telecomunicações, com o processo de internacionalizar a sua operação, contando já com funções a nível europeu. Ainda importante para o grupo, a REN é responsável pela Gestão dos Serviços de Sistema necessários à operação em segurança do sistema elétrico e opera o Mercado de Serviços de Sistema, através do qual assegura a contratação e liquidação, garantindo o equilíbrio entre a produção e o consumo de eletricidade. Para manter esta responsabilidade, a REN tem um sistema de previsão de consumo de eletricidade, que será estudado em detalhe neste projeto.

Tendo em conta um dos maiores mercados a nível nacional, os dados relacionados com a eletricidade têm de ser o mais corretos possíveis e também o mais aproximados à realidade no que toca à previsão de consumo. Para tal têm de ser considerados fatores que influenciam o consumo de eletricidade de modo direto ou indireto, como por exemplo, fatores meteorológicos, fatores de calendário, fatores económicos, sociais, entre outros.

Os dois fatores mais importantes, aqueles que influenciam diretamente o consumo de eletricidade são o fator meteorológico, mais especificamente a temperatura e o fator calendário, em particular, os dias da semana e feriados.

Dependendo se está frio ou calor, as pessoas e até os escritórios de empresas utilizam o ar condicionado, sendo este o exemplo indicado do efeito que a temperatura tem sobre a sociedade podendo recorrer a mais ou menos energia para se sentir confortável no seu dia-a-dia. O efeito calendário influencia na diferença que existe na rotina das pessoas nos dias da semana, principalmente nos fins-de-semana e feriados. Ao mudar essa rotina, muda o consumo de eletricidade em casa e nos escritórios, pois as pessoas não se deslocaram para o local de trabalho diário.

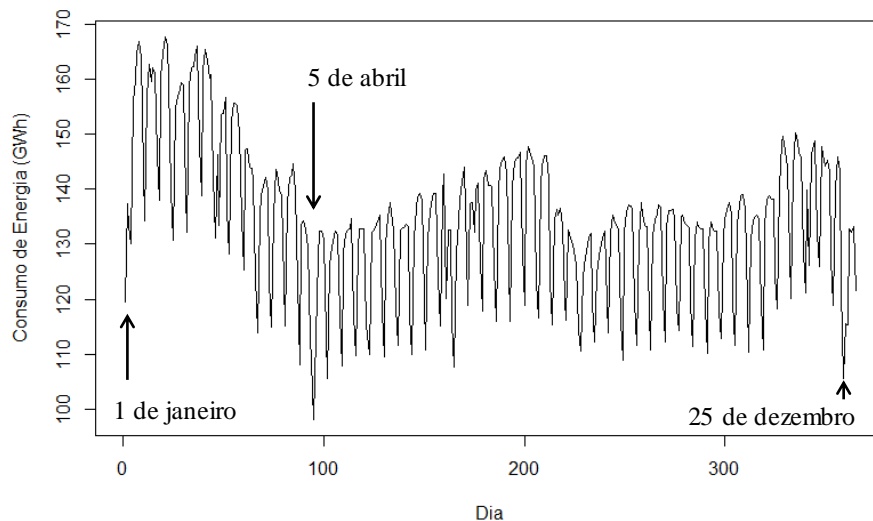


Figura 1.1 – Consumo de eletricidade em Portugal do ano 2015.

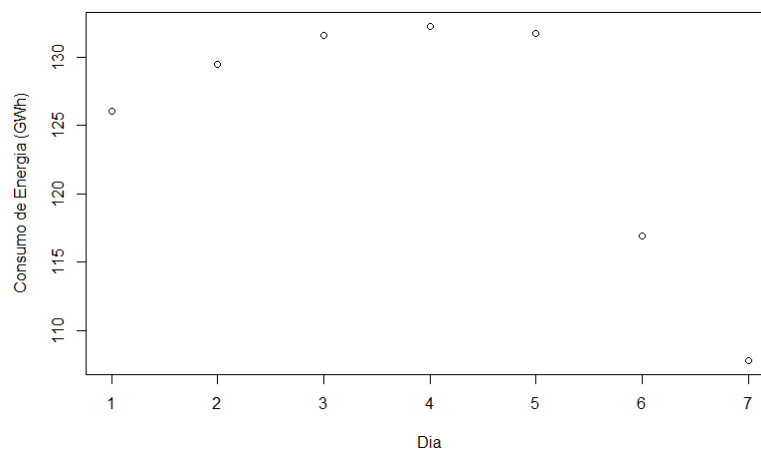


Figura 1.2 – Padrão semanal de consumo de eletricidade em Portugal.

Através das figuras 1.1 e 1.2, podemos ver essas influências, em que no primeiro gráfico temos o consumo de eletricidade do ano 2015, onde se pode observar que os dias com um consumo relativamente baixo, correspondem aos grandes feriados nacionais – 1 de janeiro, corresponde ao Ano Novo, 5 de abril, ao feriado da Páscoa e 25 de dezembro, ao Natal. Também se consegue constatar que nos meses de janeiro e fevereiro (primeiros valores do gráfico), o consumo é muito mais alto que os restantes meses, pois coincidem com os meses mais frios do ano.

Já no segundo gráfico está presente o padrão de consumo de eletricidade semanal (segunda-feira a domingo), no qual se verifica que os dias com menor consumo são sábado e domingo, isto é, os fins-de-semana.

II. Objetivos

Este projeto tem diferentes objetivos que serão atingidos por várias fases distintas. No geral, o que se pretende com este projeto é estudar e analisar a previsão de consumo de eletricidade e de gás em Portugal, de forma a melhorar o modelo de previsão utilizado pela REN, tentando obter menores erros de previsão e testando métodos inovadores através da inclusão de novas variáveis.

Atualmente o modelo utilizado por esta empresa para o consumo de eletricidade é um modelo *Holt-Winters* multiplicativo alterado. Este modelo inclui os parâmetros de nível, tendência e sazonalidade e tem como entrada a variável consumo de eletricidade corrigida por vários fatores. Como referido, estes fatores influenciam diretamente o consumo de eletricidade e, por isso, é necessária uma correção ao valor do consumo baseada na sua relação para ter uma análise correta da previsão. Estes fatores incluem a temperatura normal diária e o efeito calendário. O modelo faz previsões semanais, pois tratam-se de previsões a médio/longo prazo, e desta forma, ao contrário de previsões diárias, permite-se retirar o ruído provocado pelo efeito calendário, ou seja, é possível filtrar estes efeitos e conseguir captar melhor a sazonalidade anual. Este processo começa por corrigir o consumo pelo fator “temperatura”, seguindo a previsão pelo modelo e finalmente a correção do fator “efeito calendário”.

Numa primeira fase deste projeto será feita uma redefinição dos termos utilizados no modelo, como por exemplo o fator temperatura. Além da reparametrização destes últimos será feito um modelo auxiliar para ajudar a descobrir quais os melhores valores para a correção do consumo. Numa fase seguinte, tem-se como objetivo a inserção do fator luminosidade, à semelhança de outros fatores que contribuam para a correção do consumo, pois acredita-se que existe uma relação entre eles. Neste sentido, e ao existir uma relação, acredita-se que a luminosidade influencia o consumo diário de eletricidade, o que leva à ideia de que a inclusão deste fator no modelo irá beneficiar os resultados.

Numa fase final do projeto, pretende-se verificar se existe uma relação entre os fatores e consumo de gás em Portugal, de modo a poder aplicar o modelo estudado no consumo da eletricidade. Tendo a confirmação positiva de que os fatores temperatura e efeito calendário influenciam o consumo de gás, procede-se à adaptação deste modelo para a sua previsão. O objetivo passa pela compreensão do modelo antes da inclusão do fator luminosidade, ou seja, o modelo antigo, mas já com a reparametrização dos termos. De forma a garantir qual o modelo mais correto para os dados do consumo de gás, irão ser feitos vários testes.

III. Estrutura do documento

No presente capítulo é feita uma introdução ao tema da energia, falando sobre as energias não renováveis e renováveis. Também se aborda um pouco o seu papel em Portugal, de forma a contextualizar a importância da energia na sociedade. Sendo esta uma grande indústria, onde se inclui a empresa REN, no qual também se explica as suas principais funções, que engloba a operação do Mercado de Serviços de sistemas, é necessário que as previsões de consumo feitas por esta empresa, sejam executadas com precisão e com o menor erro possível, incluindo as variáveis mais relevantes que o possam influenciar.

No capítulo dois é introduzido o modelo utilizado pela empresa. É feita uma apresentação do modelo teórico, seguido da explicação de todo o processo que envolve a variável consumo, isto é, todo o processo de correção da variável dos fatores temperatura e efeito calendário, ambos referidos anteriormente, e do modelo modificado e os respetivos parâmetros.

O capítulo três é dedicado à metodologia utilizada no projeto. Sabendo que o modelo principal de previsão já existia, tendo sido adaptado e utilizado pela empresa, os objetivos deste projeto não se concentram diretamente no modelo, mas sim no processo à volta deste. Para certos objetivos foram utilizadas as mesmas metodologias.

Finalmente, no capítulo quatro, são apresentados os resultados e respectivas interpretações, sendo que para cada parte distinta do projeto serão apontadas as análises, os testes realizados e a discussão dos resultados obtidos. No final do documento, será feito um resumo em jeito de conclusão do projeto com os resultados principais de cada fase para verificar se os objetivos foram atingidos e para retirar as principais ideias de cada uma, de forma a produzir um modelo apropriado e eficaz para a empresa.

Capítulo 2 – Método de Previsão

I. Séries Temporais

Uma série temporal, por vezes denominada por sucessão cronológica, é um conjunto de observações registadas sequencialmente ao longo do tempo, ou seja, uma realização de um processo estocástico. Normalmente a notação utilizada é: $\{X_t, t \in T\}$ (quando T é um espaço discreto) ou $\{X(t), t \in T\}$ (quando T é um espaço contínuo), sendo T o espaço de índices e toma valores num conjunto S , o espaço de estados.

Existem diversos tipos de processos estocásticos, nos quais se incluem o univariado e o multivariado, dependendo da dimensão de T , $\dim(T) = 1$ ou $\dim(T) > 1$, respetivamente. Os processos estocásticos também podem ser unidimensionais ou multidimensionais consoante a dimensão de S ($\dim(S) = 1$ ou $\dim(S) > 1$). As observações do processo estocástico podem ser regular ou irregularmente espaçadas, dependendo se estas são feitas em intervalos de tempo regulares, ou não.

As séries temporais poderão deter algumas componentes, como a tendência e a sazonalidade. A tendência é a inclinação, positiva ou negativa, que a série apresenta, ou seja, quando o conjunto de observações da série mostra uma variação em termos de valor médio que deverá ser do mesmo sinal. A sazonalidade é a presença de uma componente periódica, isto é, variações de natureza cíclica.

Para analisar estas características existem vários métodos, sendo o mais utilizado o periodograma. O periodograma é uma função que nos mostra se a série apresenta tendência e quais os seus períodos de sazonalidade, se existirem; ou seja, a sua principal função é procurar as componentes periódicas da série temporal, através da análise das frequências de Fourier. O gráfico obtido pela função mostra essas frequências (eixo das abcissas) associadas à respetiva amplitude (eixo das ordenadas) que procura estimar a contribuição dessa frequência para a série. As frequências que tiverem associadas a uma ordenada maior, ou seja, a um “pico”, correspondem a períodos presentes na série.

II. Modelos de Previsão

A modelação de séries temporais é tipicamente feita através de modelos paramétricos estocásticos discretos, pois muitas das séries são séries temporais univariadas discretas.

Os métodos de previsão podem ser classificados como métodos quantitativos e métodos qualitativos e a sua ideia baseia-se na extrapolação das características das observações passadas e no seu inter-relacionamento. Isto implica o pressuposto de que as observações passadas contêm informações sobre o padrão de comportamento da série e que este é recorrente no tempo. Pretende-se distinguir esse padrão de qualquer ruído que possa existir e usar esse padrão para previsões para períodos de tempo subsequentes. Os modelos de suavização assumem que os valores extremos da série representam aleatoriedade e através da suavização desses extremos pode-se identificar o padrão e usá-lo para respetivas previsões. Estes são modelos paramétricos, pois trata-se de uma análise no domínio do tempo.

Os modelos mais simples incluem o *Naive* que, consiste na previsão ser igual à última observação ($\hat{y}_{k+p|k} = y_k$) e média móvel $MA(q)$, em que a previsão é uma média ponderada das q observações

passadas. Também existe o modelo AR (Auto-regressivo) de ordem p , onde a previsão depende das p anteriores, e os modelos ARMA (Auto-regressivo de médias móveis), uma extensão do anterior, em que é incluído uma componente de modelagem do erro, que é uma combinação linear destes. Por fim, os métodos de alisamento exponencial que se estendem ao modelo de Holt e ao modelo de Holt-Winters.

Para estes modelos existem medidas de desempenho preditivo que ajudam a medir a precisão e a eficácia dos resultados obtidos através dos diferentes métodos de previsão. Estas baseiam-se no cálculo do erro médio de previsão de cada um dos períodos, independentemente do sentido (positivo ou negativo) da flutuação. Algumas das medidas de desempenho preditivo mais conhecidas são as seguintes:

- MAE (*Mean absolute error*):

$$\frac{1}{N-n} \sum_{k=n+1}^N |e_{k|k-p}| \quad (2.1)$$

- RMSE (*Root mean squared error*):

$$\sqrt{\frac{1}{N-n} \sum_{k=n+1}^N e_{k|k-p}^2} \quad (2.2)$$

- MAPE (*Mean absolute percentage error*):

$$\frac{100}{N-n} \sum_{k=n+1}^N \frac{|e_{k|k-p}|}{y_k}, \quad \text{supondo } y_k > 0, \forall k \quad (2.3)$$

onde $\{y_k\}_{k=1,\dots,N}$ é a série temporal univariada unidimensional, $\hat{y}_{k+p|k}$ é a previsão do valor y_{k+p} , com origem k e horizonte p (ou seja, previsão com base na informação até k e “ p passos à frente”). O erro (empírico) da estimação é dado pela expressão $e_{k|j} = y_k - \hat{y}_{k|j}$, em particular $e_{k+1|k}$ é o erro de previsão um passo à frente. Quanto menor for o resultado da medida de desempenho, melhor é considerado o método de previsão.

a. Modelo de Alisamento Exponencial Simples

Os métodos de alisamento exponencial são métodos de filtragem e previsão baseados na estimação recursiva de parâmetros variáveis no tempo. Uma estimação condicionada por hiperparâmetros (constantes) que determinam a rapidez da variação dos parâmetros em reação às variações da série. O modelo de alisamento exponencial atribui pesos que decrescem exponencialmente de acordo com as observações mais antigas, isto é, as mais recentes têm maior peso na previsão, no entanto, todas as observações anteriores contribuem. Através das medidas de desempenho pretende-se otimizar os hiperparâmetros.

O modelo sendo o mais simples, é usado apenas para séries sem as características tendência e sazonalidade, onde se estima um parâmetro de nível da série.

Estimação dos parâmetros:

$$a_k = \alpha y_k + (1 - \alpha) a_{k-1}, \quad 0 < \alpha \leq 1 \quad (2.4)$$

Estimação de observações:

$$\hat{y}_{k+p|k} = a_k, \quad 0 \leq k \leq N; p \geq 0 \quad (2.5)$$

b. Modelo de Holt

O modelo de Holt é uma extensão do modelo de alisamento exponencial, onde é incluído um novo parâmetro relativo à tendência. O método baseia-se na estimação de 2 parâmetros, a_k (nível) e b_k (declive), e usa 2 hiperparâmetros.

Estimação dos parâmetros:

$$a_k = \alpha y_k + (1 - \alpha)(a_{k-1} + b_{k-1}), \quad 0 < \alpha \leq 1 \quad (2.6)$$

$$b_k = \beta(a_k - a_{k-1}) + (1 - \beta)b_{k-1}, \quad 0 < \beta \leq 1 \quad (2.7)$$

Estimação de observações:

$$\hat{y}_{k+p|k} = a_k + p b_k, \quad 0 \leq k \leq N; p \geq 0 \quad (2.8)$$

c. Modelo de Holt-Winters

O modelo Holt-Winters é um método de alisamento exponencial, extensão do modelo de Holt, para aplicação a séries com tendência localmente linear e efeitos de sazonalidade (de período m). De forma a analisar as novas componentes são introduzidos novos parâmetros, c_k , de forma a dessazonalizar as observações, permitindo corrigir as estimativas da tendência que resultariam do método Holt.

Existem 2 versões, diferenciando a sazonalidade:

- Sazonalidade tipo multiplicativo ($Y=T \times S+E$);

Estimação de parâmetros:

$$a_k = \alpha \left(\frac{y_k}{c_{k-m}} \right) + (1 - \alpha)(a_{k-1} + b_{k-1}), \quad 0 < \alpha \leq 1 \quad (2.9)$$

$$b_k = \beta(a_k - a_{k-1}) + (1 - \beta)b_{k-1}, \quad 0 < \beta \leq 1 \quad (2.10)$$

$$c_k = \gamma \left(\frac{y_k}{a_k} \right) + (1 - \gamma)c_{k-m}, \quad 0 < \gamma \leq 1 \quad (2.11)$$

Estimação de observações:

$$\hat{y}_{k+p|k} = (a_k + p b_k) c_j, \quad 0 \leq k \leq N; p \geq 0 \quad (2.12)$$

$$j = k + p - \left\lfloor \frac{p}{m} \right\rfloor \times m \quad (2.13)$$

Os parâmetros c_k representam fatores de correção sazonal.

- Sazonalidade tipo aditivo ($Y=T+S+E$).

Estimação de parâmetros:

$$a_k = \alpha(y_k - c_{k-m}) + (1 - \alpha)(a_{k-1} + b_{k-1}), \quad 0 < \alpha < 1 \quad (2.14)$$

$$b_k = \beta(a_k - a_{k-1}) + (1 - \beta)b_{k-1}, \quad 0 < \beta < 1 \quad (2.15)$$

$$c_k = \gamma(y_k - a_k) + (1 - \gamma)c_{k-m}, \quad 0 < \gamma < 1 \quad (2.16)$$

Estimação de observações:

$$\hat{y}_{k+p|k} = (a_k + pb_k) + c_j, \quad 0 \leq k \leq N; p \geq 0 \quad (2.17)$$

$$j = k + p - \left\lfloor \frac{p}{m} \right\rfloor \times m \quad (2.18)$$

Os parâmetros c_k representam termos de correção sazonal.

A diferença entre estes dois tipos de sazonalidade verifica-se através de uma análise gráfica. O modelo de sazonalidade multiplicativa é aplicado quando se observa que quando a tendência aumenta (ou decresce), a variabilidade das observações também aumenta (ou decresce), como se vê na figura 2.1.

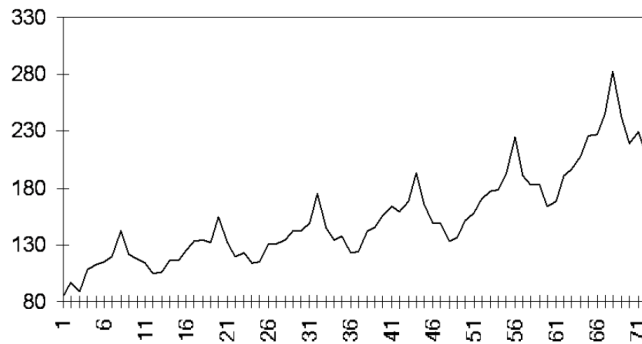


Figura 2.1 – Exemplo sazonalidade multiplicativa.

d. Tipos de Tendência

Tal como referido, o modelo de alisamento é um método de previsão onde não se verifica tendência e sazonalidade. No caso de presença de tendência, existe a extensão deste método, o modelo Holt, e na presença de sazonalidade, o modelo Holt-Winters. Assim como existem dois tipos de sazonalidade, multiplicativa e aditiva, também podem existir vários tipos de tendência. Combinando os termos relativos à tendência, a (nível) e b (declive), é possível obter cinco tipos de tendências futuras. Se T_h representar a tendência da previsão ao longo do período de tempo h e ϕ representar o parâmetro de amortecimento ($0 < \phi < 1$), tem-se os seguintes tipos de padrões de tendência:

- Nenhum: $T_h = a$;
- Aditivo: $T_h = a + bh$;
- Aditivo amortecido: $T_h = a + (\phi + \phi^2 + \dots + \phi^h)b$;
- Multiplicativo: $T_h = ab^h$;
- Multiplicativo amortecido: $T_h = ab^{(\phi + \phi^2 + \dots + \phi^h)}$.

O método da tendência amortecida é apropriado quando existe uma tendência que só se dá num curto espaço de tempo, particularmente para séries com horizontes de tempo muito longos. A tendência multiplicativa verifica-se quando a tendência não é linear, ou seja, a cada observação o declive da tendência pode variar.

III. Processo de previsão da empresa¹

O processo de previsão da empresa REN – Redes Energéticas Nacionais passa por vários passos de correção da série de consumo até à previsão propriamente dita pelo modelo Holt-Winters. A série estudada pela empresa é o consumo de eletricidade em Portugal. Como medida de desempenho é utilizado o erro em percentagem, fórmula 2.19, mas ao longo do projeto irá ser utilizado o RMSE.

$$\frac{\text{Consumo Verificado} - \text{Consumo Previsto}}{\text{Consumo Verificado}} \times 100 \quad (2.19)$$

O primeiro passo é a correção da temperatura. Como já foi abordado, a temperatura tem bastante peso no consumo diário da sociedade. Este processo é o mais complicado, pois é constituído por uma fórmula que envolve os desvios de temperatura em relação à temperatura normal e os gradientes obtidos da relação não linear entre os desvios percentuais de consumo e desvios da temperatura. No passo seguinte é feito o consumo médio semanal em dia útil, obtendo uma série semanal, para a aplicação do modelo de previsão (semanas fixas, ver anexos, pp. 55). Por fim é aplicada a correção do efeito calendário, onde são identificados todos os tipos de dias, desagregando as previsões semanais em previsões diárias. Opcionalmente, poder-se-á aplicar de forma inversa a correção da temperatura, no caso de se ter uma previsão desta que possa influenciar o consumo futuramente.

Todo este processo é feito computacionalmente, na linha de comandos, através de vários ficheiros ASCII e ficheiros executáveis, onde parte da informação é carregada a partir de uma base de dados (anexos, pp. 56).

a. Série de consumo de eletricidade

A série de consumo de eletricidade é uma série diária com dados relativos ao país inteiro. Na figura 2.2 pode-se ver o consumo de eletricidade nos anos de 2006 até 2015, ou seja, 10 anos de consumo. Na figura 2.3 tem-se a série de consumo de eletricidade no ano de 2015 apenas. É uma série univariada discreta e unidimensional, o tipo mais comum neste tipo de análise.

Observando o comportamento da série não é visível tendência ao longo dos anos, mas com uma análise mais profunda, esta existe. No entanto, a presença de sazonalidade é óbvia, com diferentes períodos. Os períodos com mais significância poderão ser o período anual e o semanal.

¹ Nota de leitura: Todas as informações neste subcapítulo dizem respeito a decisões da empresa e só da empresa REN – Redes Energéticas Nacionais

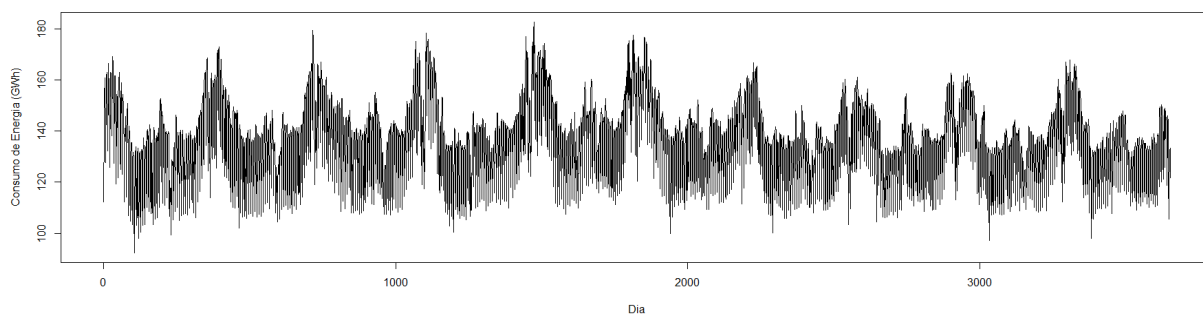


Figura 2.2 – Série de consumo de eletricidade nos anos 2006 até 2015.

Na série relativa apenas ao ano 2015, nota-se a presença de tendência, algo irregular, e mais uma vez a existência de sazonalidade.

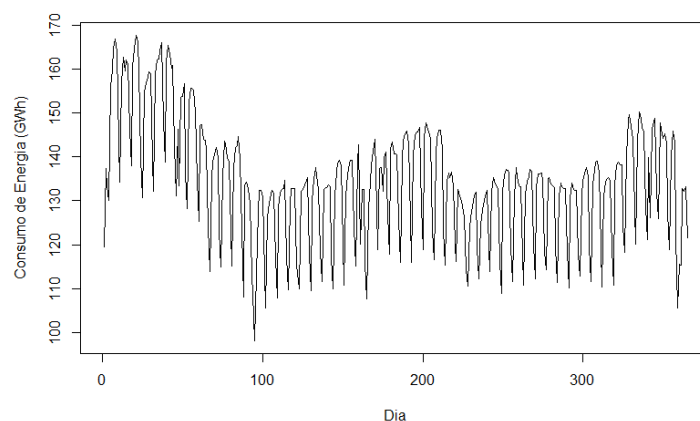


Figura 2.3 – Série de consumo de eletricidade do ano 2015.

As figuras 2.4 e 2.5, apresentam os periodogramas das séries anteriores. O periodograma da série completa, com o consumo dos anos de 2006 até 2015, mostra-nos mais do que 4 “picos”, tendo alguns com uma frequência baixa, isto é, um período grande. Os “picos” mais altos correspondem a períodos de 365 dias e 7 dias, como se tinha dito, os períodos mais significativos.

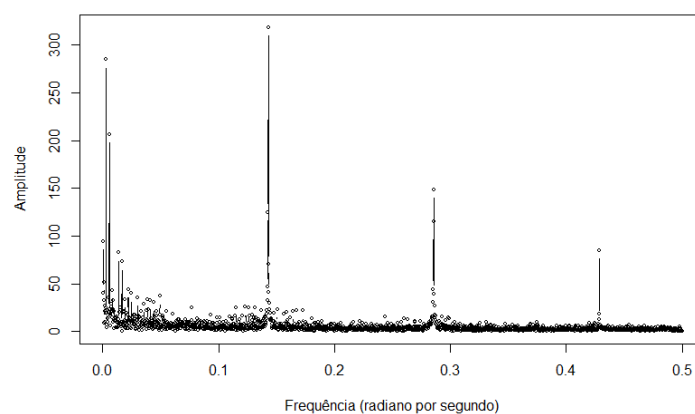


Figura 2.4 – Periodograma da série referente ao consumo de eletricidade em Portugal nos anos 2006 até 2015.

Na figura 2.5, no periodograma da série referente ao consumo de eletricidade em Portugal no ano de 2015 existem quatro “picos”, o que significa que existem quatro períodos de sazonalidade, onde o maior, que é o mais relevante para este projeto, corresponde a um período de 7 dias.

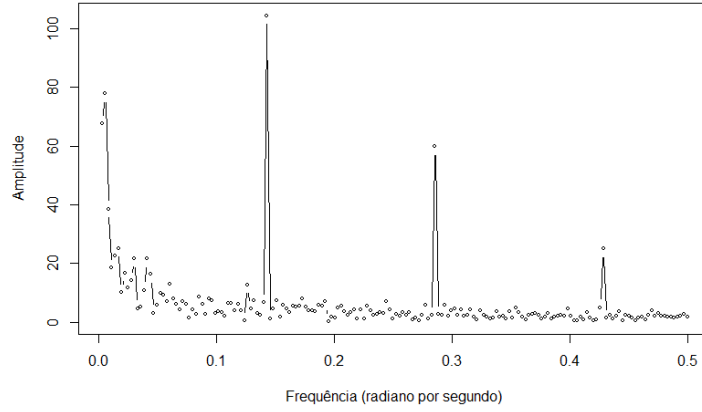


Figura 2.5 – Periodograma da série referente ao consumo de eletricidade em Portugal no ano de 2015.

Apesar de não se notar tendência na série mais completa, esta existe, mesmo que tenha alguma irregularidade. Além desta, também existe uma forte componente de sazonalidade, pelo que o modelo adaptado, o modelo Holt-Winters é o mais indicado para previsão desta série.

b. Correção Temperatura

Seja C_{ikj} o consumo do dia i da semana k do ano j . Neste processo, calcula-se o consumo referido à temperatura normal C_{ikj}^N , isto é, o consumo que se registaria caso a temperatura ocorrida nesse dia tivesse sido a temperatura considerada *normal* e não a que se verificou. Para tal calcula-se a temperatura normal, que consiste no ajustamento polinomial da série do cálculo do valor médio dos valores médios diários, para cada dia diferente do ano, desde 1977 (temperaturas registadas em Lisboa). Os valores médios diários são a média entre a temperatura mínima e a temperatura máxima.

Inicialmente este processo de correção ao consumo era realizado através da fórmula:

$$C_{ikj}^N = \frac{C_{ikj}}{1 + \gamma_k}, \quad (2.20)$$

onde γ_k é o parâmetro de regressão que traduz a sensibilidade do consumo aos desvios da temperatura atmosférica relativamente ao seu valor médio, ou seja, um gradiente para cada uma das 52 semanas do ano, que refletia uma relação linear entre as variáveis consumo e temperatura.

À série de desvios de temperatura efetua-se um alisamento exponencial, de modo a suavizar as mudanças bruscas de temperatura:

$$\phi_{ikj} = \lambda(\theta_{ikj} - \bar{\theta}_{ikj}) + (1 - \lambda)\phi_{(i-1)kj} \quad (2.21)$$

$\bar{\theta}_{ikj}$ – temperatura média normal utilizando o histórico a partir do ano 1977

θ_{ikj} – média aritmética das temperaturas máxima e mínima ocorridas no dia ikj

λ – parâmetro que pretende representar o fenómeno físico de inércia térmica, $0 < \lambda < 1$

Atualmente $\lambda = 0,24$, cujo valor foi estimado utilizando o modelo Holt-Winters aplicado a dados de anos passados de forma a minimizar a variância dos erros de previsão e_{ikj} .

Mas rapidamente, devido à ocorrência de temperaturas extremas, levou-se a postular a hipótese de uma relação não linear entre os desvios percentuais do consumo de energia e os desvios absolutos de temperatura verificada em relação à temperatura *normal*. Através de uma amostra alargada de valores históricos foi possível tornar esta hipótese viável.

Deste modo, deixa-se de ter apenas um gradiente γ_k e passa-se a ter três valores distintos. Estes três gradientes continuam a ser a relação linear entre os desvios percentuais de consumo e os desvios absolutos de temperatura (resultados do anterior alisamento exponencial), mas em intervalos separados pelos valores g_{inf} e g_{sup} , parâmetros semanais podendo tomar valores diferentes para cada uma das 52 semanas, com a imposição da continuidade desses três gradientes (g_{1k} , g_{2k} e g_{3k}) nos pontos g_{inf} e g_{sup} . No entanto, devido ao desenvolvimento necessário para calcular estes dois últimos valores para as 52 semanas, chegou-se à conclusão que este não seria recompensado com os resultados, ficando-se então com:

$$g_{inf} = -2 \text{ e } g_{sup} = 2, \forall \text{ semana } k$$

Os valores escolhidos para estes dois pontos devem-se ao desvio-padrão dos desvios de temperatura de cada um dos 365 dias que ronda o valor 2, podendo ultrapassar o valor 2,5 (ver anexos, pp.57). A fórmula utilizada foi:

$$\sigma_i = \sqrt{\sum_j \frac{(Tmed_{ij} - Tnor_{ij})^2}{n^o \text{ de anos}}} \quad (2.22)$$

$Tmed_{ij}$ – temperatura média do dia i do ano j

$Tnor_{ij}$ – temperatura *normal* do dia i do ano j

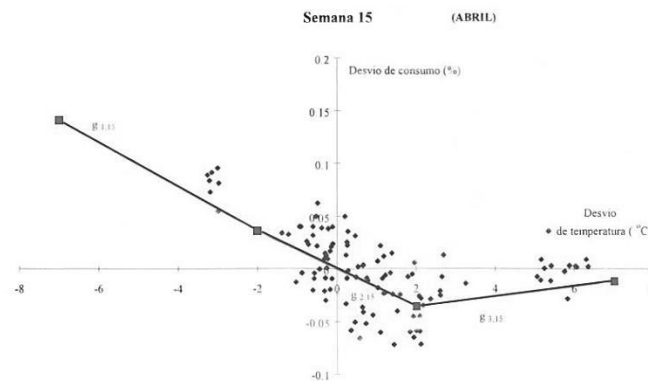


Figura 2.6 – Relação entre os desvios de consumo e desvios de temperatura na semana 15.

Como se pode ver na figura 2.6 depreende-se a existência de três retas com declives diferentes, divididas pelos pontos -2 e 2.

Este parâmetro é estimado utilizando o próprio modelo de Holt-Winters, por regressão linear entre os erros relativos da previsão de consumo em relação à temperatura normal e os erros da temperatura, obtidos do alisamento exponencial.

Por fim, obtém-se o consumo referido à temperatura normal através do seguinte cálculo:

$$C_{ikj}^N = \frac{C_{ikj}}{1 + G_{ikj}} = C_{ikj} * \frac{1}{1 + G_{ikj}} \quad (2.23)$$

Onde,

$$G_{ikj} = g_{1k} * \text{Min}(0, \phi_{ikj} - g_{\text{inf}}) + g_{2k} * \text{Max}(\text{Min}(\phi_{ikj}, g_{\text{sup}}), g_{\text{inf}}) + g_{3k} * \text{Max}(0, \phi_{ikj} - g_{\text{sup}}) \quad (2.24)$$

A formula anterior condensa o gráfico de uma função linear por troços, que se pode ver no exemplo a seguir:

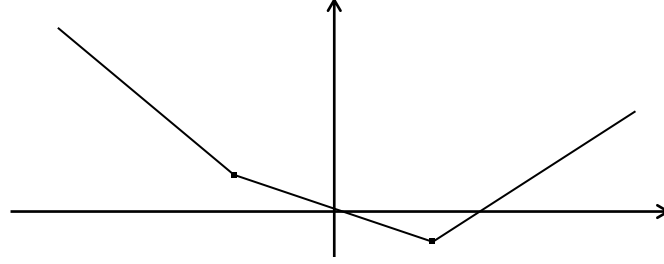


Figura 2.7 – Exemplo três gradientes.

- Se $\phi_{ikj} < -2$ (gradiente 1),

$$G_{ikj} = g_{1k} * (\phi_{ikj} - (-2)) + g_{2k} * (-2) \quad (2.25)$$

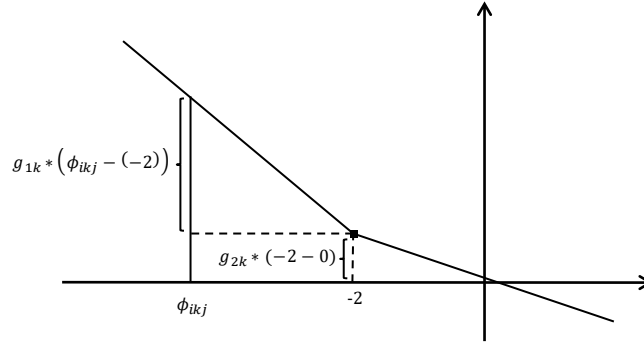


Figura 2.8 – Exemplo caso $\phi_{ikj} < -2$

- Se $-2 \leq \phi_{ikj} < 2$ (gradiente 2),

$$G_{ikj} = g_{2k} * (\phi_{ikj}) \quad (2.26)$$

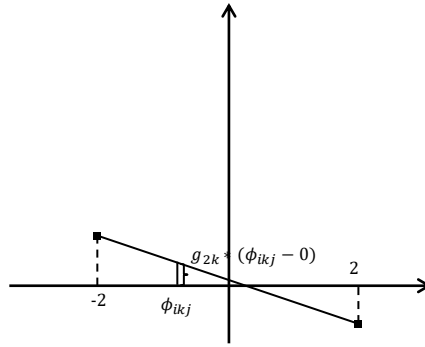


Figura 2.9 – Exemplo caso $-2 \leq \phi_{ikj} < 2$

- Se $\phi_{ikj} \geq 2$ (gradiente 3),

$$G_{ikj} = g_{2k} * (2) + g_{3k} * (\phi_{ikj} - 2) \quad (2.27)$$

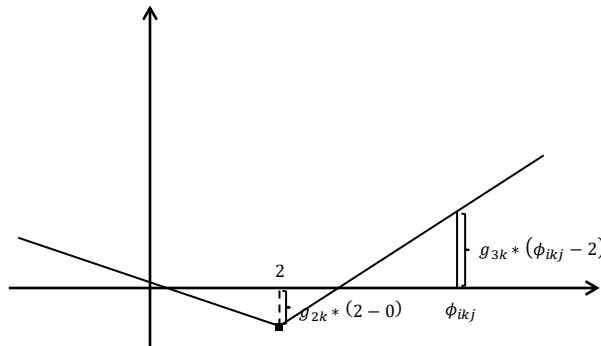


Figura 2.10 - Exemplo caso $\phi_{ikj} \geq 2$

Para cada intervalo de desvios de temperatura, a fórmula G_{ikj} difere, como se acabou de verificar neste exemplo, influenciando o fator de correção de acordo com estes.

c. Modelo

Tal como abordado no início deste capítulo, o modelo utilizado pela empresa é um modelo de Holt-Winters de sazonalidade e tendência multiplicativa, o que levou a algumas alterações na estimação de observações e de parâmetros.

Como o modelo faz previsões semanais, revelou-se necessário que a série de entrada seja também semanal. Para tal, antes da previsão é feita uma agregação. Essa agregação consiste em calcular o consumo médio semanal em dia útil, ou seja, o consumo médio dos dias úteis de cada semana.

$$\bar{C}_{kj}^N = \frac{\sum_i C_{ikj}^N}{n^{\circ} \text{ de dias \u00fateis da semana } k} \quad (2.28)$$

A série utilizada no modelo é a série de consumos médios semanais em dia útil, \bar{C}_{kj}^N (conhecida até à semana k do ano j), no modelo designada por X_T , para facilitar a nomenclatura. Os primeiros valores históricos conhecidos foram utilizados para estimar os valores iniciais dos parâmetros.

Estimação dos parâmetros:

$$a_T = \alpha \left(\frac{X_T}{c_{T-m}} \right) + (1 - \alpha) \times a_{T-1} \times b_{T-1}, \quad 0 \leq \alpha \leq 1 \quad (2.29)$$

$$b_T = \beta \left(\frac{a_T}{a_{T-1}} \right) + (1 - \beta) b_{T-1}, \quad 0 \leq \beta \leq 1 \quad (2.30)$$

$$c_T = \gamma \left(\frac{X_T}{a_T} \right) + (1 - \gamma) c_{T-m}, \quad 0 \leq \gamma \leq 1 \quad (2.31)$$

Estimação das observações:

$$\hat{X}_{T+p|T} = (a_T \times b_T^p) c_{T+p|T+p-m}, \quad 0 \leq T \leq N; p \geq 0 \quad (2.32)$$

Como se pode ver as mudanças foram feitas nos parâmetros de nível e tendência, assim como na previsão, pois tratando-se de uma tendência multiplicativa, altera-se os parâmetros e a equação de previsão relativos a esta característica.

d. Efeito Calendário

Além do peso que a temperatura tem sobre o consumo de eletricidade, o efeito calendário também mostrou que influencia de forma bastante notória a vida diária da sociedade que, no contexto deste processo, passa pela aplicação da função calendário:

$$\psi(ikj) = td, \quad (2.33)$$

sendo td , a classificação dos tipos de dia. Esta classificação começou com a definição de 10 classes de tipo de dia:

- **FN** – Feriado Nacional;
- **DE** – Dia Especial;
- **DN** – Domingo Normal;
- **SE** – Sábado Especial: sábado a seguir a feriado;
- **SN** – Sábado Normal;
- **P6** – Sexta-feira quando quinta-feira foi feriado;
- **P2** – Outras pontes: segunda-feira quando terça-feira é feriado, terça, quarta ou quinta-feira situadas entre feriados;
- **CA** – Dias de cais: segunda, terça, quarta ou quinta-feira quando a véspera foi feriado
- **F2** – Segunda-feira normal;
- **DU** – Dia útil: dia não incluído nas classes anteriores.

Para além destas classes, os dias são diferenciados consoante a semana fixa a que pertencem, e agrupados por alguns comportamentos idênticos. Assim resultam 511 tipos de dia.

A previsão de consumo para cada dia, tendo em conta os efeitos de calendário, é dada por:

$$\hat{C}_{ikj}^N = r(\psi(ikj)) \cdot \hat{C}_{kj}^N \quad (2.34)$$

Onde $r(td)$ é o valor esperado do quociente entre o consumo do dia classificado como td pela função $\psi(.)$ e o valor do consumo médio em dia útil dessa semana. $r(td)$ é um parâmetro que representa o peso do consumo de cada dia relativamente a um dia útil da mesma semana, tendo para um $r(\text{dia útil}) = 1$.

e. Efeito da Temperatura

Opcionalmente é possível aplicar o inverso da correção da temperatura, de forma a obter uma previsão influenciada por uma previsão de temperatura:

$$\hat{\hat{C}}_{ikj} = \hat{C}_{kj}^N (1 + G_{ikj}) \quad (2.35)$$

Onde

$$\begin{aligned} G_{ikj} = & g_{1k} * \text{Min}(0, \hat{\phi}_{ikj} - g_{inf}) + \\ & g_{2k} * \text{Max}(\text{Min}(\hat{\phi}_{ikj}, g_{sup}), g_{inf}) + \\ & g_{3k} * \text{Max}(0, \hat{\phi}_{ikj} - g_{sup}) \end{aligned} \quad (2.36)$$

e

$$\hat{\phi}_{ikj} = \lambda(\hat{\theta}_{ikj} - \bar{\theta}_{ikj}) + (1 - \lambda)\phi_{(i-1)kj} \quad (2.37)$$

é o valor previsto da variável ϕ_{ikj} obtida a partir do alisamento exponencial 2.20 feito na correção de temperatura, utilizando as temperaturas conhecidas até ao instante inicial de previsão e uma eventual estimativa adicional $\hat{\theta}_{ikj}$ para os instantes seguintes.

Capítulo 3 – Metodologia

I. Coeficiente de correlação e dependência linear

O coeficiente de correlação permite avaliar a dependência linear entre duas variáveis aleatórias com segundo momento finito. O coeficiente de correlação de Pearson:

$$\rho_{X,Y} = \text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \quad (3.1)$$

é uma medida adimensional desta relação, $-1 \leq \rho_{X,Y} \leq 1$, onde -1 e 1 significam dependência linear (declive negativo e positivo respetivamente) e 0 corresponde a ausência completa desta. Note-se que dependência ou uma associação forte não significa obrigatoriamente uma relação de causa-efeito entre as variáveis. Como exemplo, na figura 3.1 apresenta-se o diagrama de dispersão da temperatura *versus* o consumo de eletricidade (2006 a 2015). Se consideramos as temperaturas abaixo de 17 graus (aproximadamente) é bem visível uma associação inversa entre as variáveis, enquanto nas temperaturas superiores os pontos apresentam uma entropia significativa.

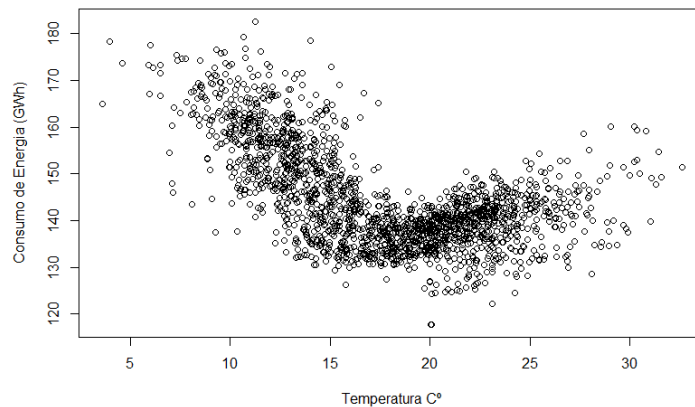


Figura 3.1 – Diagrama de dispersão entre a temperatura e o consumo de eletricidade nos anos de 2006 até 2015.

O coeficiente de correlação é estimado com base numa amostra, por:

$$r_{X,Y} = \frac{\sum_{i=1}^n (X_i - \bar{x})(Y_i - \bar{y})}{(n-1)S_x S_y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3.2)$$

Se considerarmos que as observações são independentes e têm distribuição normal bivariada é possível definir testes de hipóteses com base neste estimador para avaliar a significância de $\rho_{X,Y}$.

Na presença de coeficientes de correlação elevados faz sentido usar modelos de regressão linear, simples ou múltipla, de que se falará no próximo subcapítulo.

II. Regressão linear múltipla

A regressão linear múltipla procura estabelecer uma relação linear entre a variável resposta e as variáveis independentes, possibilitando ver o efeito conjunto destas variáveis, bem como o efeito marginal de cada característica, X_i , na variável dependente. Os coeficientes β_i indicam a alteração da

variável resposta por variação unitária do i -ésimo preditor, mantendo-se constante o nível das restantes variáveis independentes. Dado um indivíduo da população, com características $\mathbf{x} = (x_1, \dots, x_p)$, o valor associado à característica Y :

$$Y_x = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon_x \quad (3.3)$$

Com os seguintes pressupostos:

- Y_x é uma variável aleatória, dependente de um vetor não aleatório \mathbf{x} .
- $\mu_x = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ é o valor médio da variável dependente Y_x .
- ε_x é o erro aleatório e corresponde a possíveis efeitos desconhecidos, que influenciam Y_x .
- Os valores de ε para diferentes elementos da População obedecem às condições de Gauss Markov:
 1. $E(\varepsilon) = 0$
 2. $Var(\varepsilon) = \sigma^2$
 3. $E(\varepsilon_{x_i} \varepsilon_{x_j}) = 0$, para quaisquer elementos da população, $i \neq j$
- $Cov(\varepsilon) = \sigma^2 I_{n \times n}$, matriz de covariância de $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$
- Os erros ε_i são i.i.d e seguem distribuição normal: $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$

Em notação matricial, o modelo pode-se escrever da seguinte maneira:

$$Y = X\beta + \varepsilon \quad (3.4)$$

com $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ 1 & \dots & x_{np} \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$ e $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$, onde se tem presente uma população Y de dimensão n .

Para o processo de estimação dos parâmetros β_i existem dois métodos, o método dos mínimos quadrados e o método da máxima verosimilhança.

- Método dos mínimos quadrados:
Estima os parâmetros de forma a encontrar o mínimo da soma dos quadrados dos resíduos:

$$SQ(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2. \quad (3.5)$$

A equação anterior em notação matricial é:

$$SQ(\beta) = (Y - X\beta)^T(Y - X\beta), \quad (3.6)$$

E derivando em ordem dos parâmetros, obtemos o estimador:

$$\frac{\partial SQ(\beta)}{\partial \beta} = 0 \Leftrightarrow \hat{\beta} = (X^T X)^{-1} X^T Y \quad (3.7)$$

- Método da máxima verosimilhança:
A função de verosimilhança é:

$$L(\beta_o, \dots, \beta_p, \sigma^2) = \prod_{i=1}^n \phi\left(\frac{Y_i - \mu_i}{\sigma}\right) / \sigma, \quad (3.8)$$

onde ϕ é a função densidade de probabilidade da Gaussiana reduzida. De forma a maximizar a função é necessário encontrar os valores de β que maximizam:

$$\begin{aligned} \text{Log}L(\beta_o, \dots, \beta_p, \sigma^2) &= \sum_{i=1}^n (\text{Log}\phi\left(\frac{Y_i - \mu_i}{\sigma}\right) - \text{Log}(\sigma)) \\ &= -n\text{Log}(2\pi\sigma^2)/2 - \sum_{i=1}^n \frac{1}{2} \left(\frac{Y_i - \mu_i}{\sigma}\right)^2, \end{aligned} \quad (3.9)$$

e derivando, obtemos o estimador:

$$\frac{\partial L(\beta)}{\partial \beta} = 0 \Leftrightarrow \hat{\beta} = (X^T X)^{-1} X^T Y. \quad (3.10)$$

Propriedades dos estimadores:

Seja $H = X(X^T X)^{-1} X^T$, com $\hat{\mu} = X\hat{\beta} = HY$ e $e = Y - X\hat{\beta} = (I - H)Y = MY$, então tem-se:

- $E[\hat{\beta}] = \beta$ e $\text{Cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$
- $E[\hat{\mu}] = \mu$ e $\text{Cov}(\hat{\mu}) = \sigma^2 H$
- $E[e] = 0$ e $\text{Cov}(e) = \sigma^2 M$

Para uma análise mais simples, é feita uma partição da soma dos quadrados:

$$SS_{Tot} = SS_{Reg} + SS_E, \quad (3.11)$$

onde,

$$SS_{Tot} = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (3.12)$$

$$SS_{Reg} = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \quad (3.13)$$

$$SS_E = \sum_{i=1}^n e_i^2 \quad (3.14)$$

A variabilidade total é assim decomposta em duas partes: uma variabilidade devido à regressão (explicada pelo modelo) e outra residual (não explicada pelo modelo). A soma dos resíduos é 0 e a soma dos valores observados é igual à soma dos valores ajustados.

O coeficiente de determinação para a equação de regressão é definido por

$$R^2 = \frac{SS_{Reg}}{SS_{Tot}} = 1 - \frac{SS_E}{SS_{Tot}}, \quad (3.15)$$

onde este varia entre 0 e 1, e quanto mais perto de 1 indica um melhor ajustamento do modelo aos dados, ou seja, que o efeito das variáveis independentes é significativo para a variável dependente.

Se atendermos às condições de Gauss-Markov e aos pressupostos distribucionais do modelo, concluímos que:

- $b \cap NM(\beta, \sigma^2(X^T X)^{-1})$
- $\frac{SS_E}{\sigma^2} \cap \chi^2_{n-k-1}$ tal que $E(MS_E) = \sigma^2$, com $MS_E = \frac{SS_E}{n-k-1}$
- b e SS_E são independentes e, portanto, SS_E e SS_{Reg} também o são
- $\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^n e_i^2}{n-k-1} = MS_E$

Para avaliar a significância do modelo tem-se o teste F com a seguinte hipótese:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0 \text{ vs } H_1: \exists \beta_i \neq 0, i = 1, \dots, p$$

Sob H_0 ,

- $\frac{SS_{Tot}}{\sigma^2}$ tem distribuição Qui-quadrado com $n-1$ graus de liberdade.
- $\frac{SS_{Reg}}{\sigma^2}$ tem distribuição Qui-quadrado com k graus de liberdade.
- $\frac{SS_E}{\sigma^2}$ tem distribuição Qui-quadrado com $N-k-1$ graus de liberdade.
- $F = \frac{SS_{Reg}}{k} / \frac{SS_E}{n-k-1} = MS_{Reg} / MS_E \cap F_{k, n-k-1}$

Comparando o p -value com um nível de significância pré-estabelecido, avalia-se a qualidade e a significância do modelo.

Para uma análise mais pormenorizada, podem ser feitos testes a cada coeficiente, dependendo do objetivo em mente. Tem-se os possíveis testes de hipóteses bilaterais (existem versões unilaterais óbvias dos mesmos):

$$H_0: \beta_i = 0 \text{ vs } H_1: \beta_i \neq 0, i = 1, \dots, p,$$

com a regra de decisão de rejeitar a hipótese nula caso

$$\left| \frac{b_i - 0}{s_{b_i}} \right| > t_{n-2, 1-\alpha/2}, \quad (3.16)$$

em que,

$$s_{b_i}^2 = \frac{MS_E}{\sum_{j=1}^n (x_j - \bar{x})^2} = \frac{MS_E}{ns_x^2} \quad (3.17)$$

Caso se rejeite a hipótese nula, tem-se que o coeficiente é significativo para o modelo de regressão e a variável independente associada a esse coeficiente explica algum do comportamento da variável dependente.

O modelo final adotado deverá ser explicativo e parcimonioso e deverá integrar apenas os preditores mais significativos. Existem métodos de seleção de variáveis e de procura de um sub-modelo final simples, como o *forward*, o *backward* e o *stepwise*.

Capítulo 4 – Resultados

I. Definição de metodologia de reparametrização

Nesta fase tem-se como objetivo calcular a nova temperatura normal resultante do ajustamento polinomial dos valores históricos das temperaturas médias diárias de Lisboa. De seguida, quer-se recalcular os gradientes utilizados na correção da temperatura, até se atingir a convergência. Nos subcapítulos que se seguem será explicado em maior detalhe o processo envolvente para se atingir os objetivos em análise.

a. Temperatura Normal

Nesta primeira parte do projeto, pretende-se encontrar o melhor grau de ajustamento exponencial para as temperaturas diárias de Lisboa, isto é, o grau do polinómio mais adequado para as temperaturas médias diárias de forma a haver consistência da curva do polinómio com os dados, mas que não existam diferenças significativas e abruptas de um dia para o outro, como se pode ver nas temperaturas atuais, na figura 4.1. Os valores do ajustamento polinomial são os valores da temperatura normal para 366 dias, que serão utilizados no modelo. Este ajustamento polinomial foi feito de forma a ser o melhor ajustamento possível, expondo aquele que apresentava o melhor R^2 e a melhor suavidade visual.

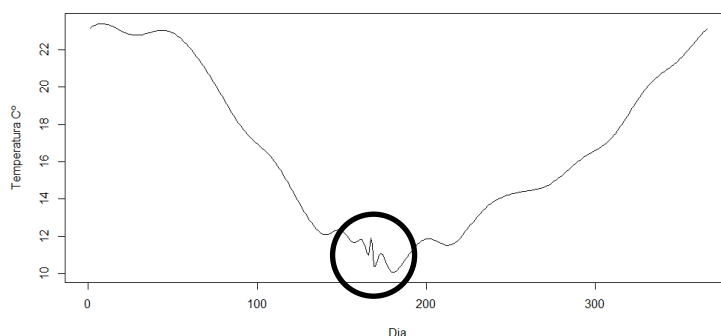


Figura 4.1 – Temperaturas normais com escala de 18 de julho a 17 de julho do ano a seguir.

Este processo inicial não envolve testar as temperaturas no modelo, *i.e.*, calcular o erro de previsão com as temperaturas obtidas. Apenas se pretende um teste independente que envolva as temperaturas mínimas e máximas registadas até à data. Desta forma, vai-se estudar qual o melhor grau do ajustamento polinomial e qual o melhor conjunto de valores históricos a utilizar.

O conjunto de valores de temperaturas normais utilizado presentemente no modelo descrito anteriormente, é o resultado do ajuste polinomial do valor médio dos valores médios diários, para cada dia diferente do ano, desde 1977. Atualmente tem-se a curva das temperaturas normais, na figura 4.2.

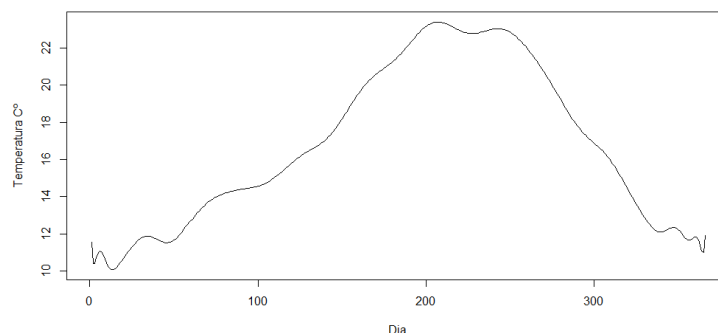


Figura 4.2 – Temperaturas normais com escala de 1 de janeiro a 31 de dezembro.

O primeiro passo começa por seleccionar as temperaturas mínimas e máximas e calcular as temperaturas médias diárias apenas de Lisboa, desde 1977 até 2015 e de seguida, a média de temperatura para cada dia do ano, de forma a comparar com as temperaturas normais utilizadas. Considera-se que a temperatura média diária é a média da temperatura mínima e temperatura máxima desse dia, pois são os dados disponíveis.

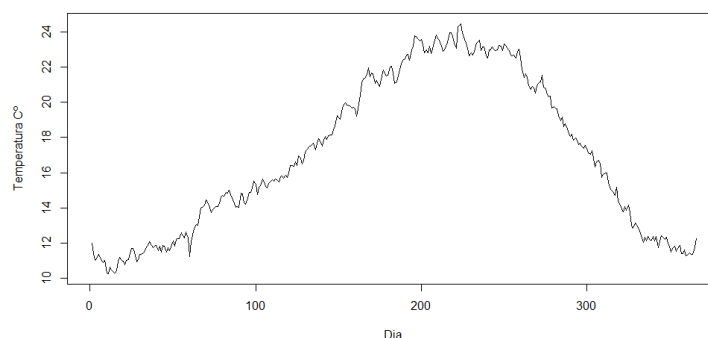


Figura 4.3 – Temperaturas médias com escala de 1 de janeiro a 31 de dezembro.

De forma a perceber como se comportam os dados, começa-se por fazer uma análise gráfica dos mesmos. Consegue-se observar na figura 4.3, que não há suavização dos dados, ou seja, as temperaturas médias de um dia para outro sobem e descem bruscamente. Portanto para um perfeito ajustamento polinomial, este teria de ser de grau bastante alto, mas como se pretende um ajustamento polinomial suave sem curvas acentuadas, procedeu-se a uma análise de regressão.

Observando o mesmo gráfico, consegue-se perceber que existem algumas subidas e descidas relevantes na curva geral dos dados, que eliminam a hipótese de se fazer uma análise de regressão linear simples, pois esta só permite que a variável independente tenha grau 1. A melhor opção para este caso é uma regressão polinomial, de forma a combinar o ajustamento polinomial com a análise de regressão linear múltipla. A regressão polinomial é o modelo de regressão linear múltipla, em que cada preditor é uma potência de x , dias.

Através do *R* com as funções *lm* e *poly*, é possível fazer-se a análise de regressão polinomial, de forma a perceber qual o grau mínimo para o ajustamento ideal aos dados. A função *lm* permite o ajustamento de modelos lineares, enquanto a função *poly* permite criar polinómios ortogonais do grau 1 até ao grau dado. Juntando estas duas funções, consegue-se obter um ajustamento polinomial.

Como explicado anteriormente, na regressão linear múltipla é realizado um teste *t-student* a cada coeficiente para verificar a sua significância para o modelo. Neste caso, é feito o teste a cada coeficiente associado a cada potência, pois a equação da regressão polinomial é dada por:

$$Y_x = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2 + \dots + \beta_p x_p^p + \varepsilon_x \quad (4.1)$$

Desta forma, é possível ver quais os graus do polinómio que podem ser significativos para o ajustamento polinomial dos dados e, visualmente, procura-se um grau que seja satisfatório e que confirme graficamente os resultados. O teste *t* é uma das informações dadas pelo *summary* da função *lm* no *R*, que do lado direito de cada coeficiente mostra os níveis de significância significativos através do símbolo ‘*’ (abaixo de 0.05%). Além da significância de cada coeficiente, avalia-se o coeficiente de determinação usual e o ajustado.

Para a descoberta do melhor grau, procedeu-se a vários testes com diferentes graus de polinómio, verificando quais são aqueles que têm significância para o modelo de regressão.

Tabela 4.1 – Significância dada pelo teste *t-student* aos coeficientes para cada ajustamento polinomial de grau 5, 10, 15 e 20.

* - existe significância

N – não existe significância

Grau Polinómio	Significância dos Coeficientes													
	≤5	6	7	8	9	10	11	12	13	14	15	16	17	≥18
5	*	-	-	-	-	-	-	-	-	-	-	-	-	-
10	*	*	*	*	N	*	-	-	-	-	-	-	-	-
15	*	*	*	*	N	*	N	*	*	*	*	-	-	-
20	*	*	*	*	N	*	N	*	*	*	*	N	*	N

Tabela 4.2 – Coeficiente de determinação (e ajustados) para cada ajustamento polinomial de grau 5, 10, 15 e 20.

Grau Polinómio	5	10	15	20
R ²	0.9882	0.9915	0.9927	0.9934
R ² ajustado	0.988	0.9913	0.9924	0.993

Na tabela 4.1, pode-se ver que, em todos os casos, os coeficientes de grau menor ou igual a 5 são significativos o que implica que para o bom ajustamento dos dados é necessário que o polinómio seja, no mínimo, de grau 5. Também se pode verificar que, para todos os casos nos quais estes se incluem, o coeficiente de grau 9 e grau 11 não tem qualquer significância para o modelo, enquanto os restantes apresentam alguma relevância para este. No caso do ajustamento com um polinómio de grau 20, percebe-se que os coeficientes mais altos não são significativos para o ajustamento, isto é, não trazem nova informação para o modelo, pelo que não compensa incluí-los no mesmo.

Comparando os coeficientes de determinação, presentes na tabela 4.2, a diferença é bastante pequena, principalmente para os ajustamentos com grau superior a 10, onde a diferença está apenas na terceira casa decimal. Assim, tal como para os coeficientes de determinação ajustados, a diferença é muito reduzida. Portanto, desta forma percebe-se que uma tomada de decisão baseada na comparação dos coeficientes de determinação não é a mais correta, pois os valores obtidos são bastantes bons e diferem pouco entre si.

Voltando à significância dos coeficientes e ao comportamento das curvas obtidas graficamente, compreende-se que, para os polinómios com grau mais baixo, a curva será demasiado suave para o objetivo que se pretende atingir, pelo que se excluem os graus abaixo de 6 ou 7. Como já se tinha visto previamente, os coeficientes com grau muito alto não apresentavam significância pelo que não se justifica os ajustamentos com grau acima de 14 ou 15, além que graficamente a curva teria oscilações indesejadas, como na curva do conjunto atual de temperaturas normais.

Por acordo mútuo com a empresa, decidiu-se estudar os polinômios com graus entre 8 e 13, estudando a respectiva curva graficamente, para obter assim o polinômio ideal. Neste intervalo de graus, exclui-se o 9 e 11 pois, em todos os testes, estes não apresentaram qualquer relevância. Então, para graus possíveis, e para o ajustamento polinomial ideal tem-se como referência os graus 8, 10, 12 e 13, com os seguintes coeficientes de determinação na tabela 4.3:

Tabela 4.3 - Coeficiente de determinação (e ajustados) para cada ajustamento polinomial de grau 8, 10, 12 e 13.

Grau Polinômio	8	10	12	13
R^2	0.9912	0.9915	0.9918	0.9924
R^2 ajustado	0.991	0.9913	0.9915	0.9921

Como a diferença entre os coeficientes de determinação e entre os coeficientes de determinação ajustados é muito baixa, passou-se à análise gráfica, para avaliar a curva correspondente.

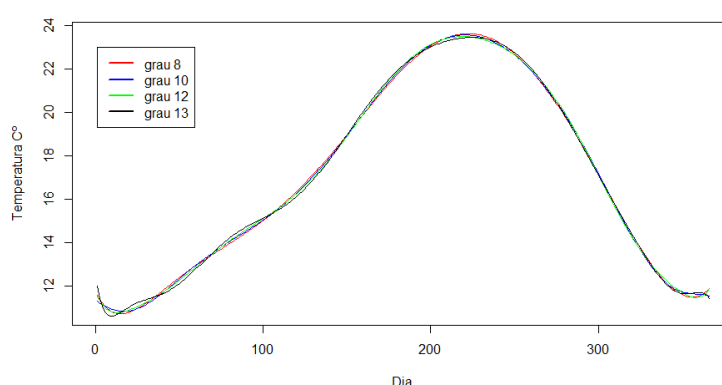


Figura 4.4 – Ajustamentos polinomiais de grau 8, 10, 12 e 13.

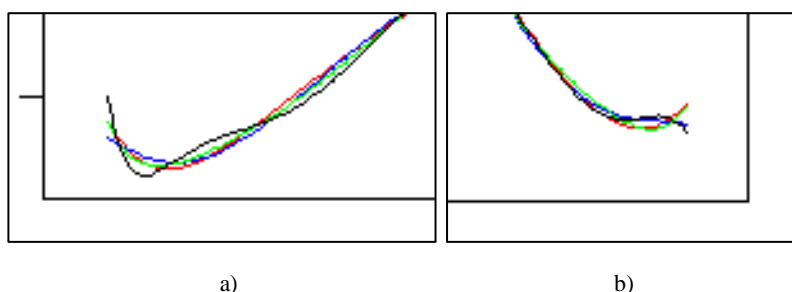


Figura 4.5 – a) Pontas iniciais das curvas dos ajustamentos polinomiais de grau 8, 10, 12 e 13.
b) Pontas finais das curvas dos ajustamentos polinomiais de grau 8, 10, 12 e 13.

Analisando o gráfico das curvas de ajustamento dos 4 casos (na figura 4.4), vê-se que no geral o comportamento destas é muito semelhante, tendo o ponto mais alto na altura do verão e as temperaturas mais baixas nas pontas dos gráficos, ou seja, no inverno. Neste sentido, é nas extremidades onde se encontram as maiores diferenças entre as curvas para poder fazer uma comparação entre a suavidade das curvas.

Na figura 4.5, tem-se uma visão aproximada das pontas e pode-se observar que a curva preta, correspondente ao ajustamento de grau 13, tem o comportamento que não se pretende, ou seja, a ponta inicial tem uma descida abrupta, seguida por algumas oscilações, e o início da curva não corresponde minimamente à ponta final. Nas restantes três curvas, já não se observa este comportamento variado de forma tão exagerada.

Tendo em conta as curvas dos restantes ajustamentos, observa-se na figura 4.6 que nas pontas iniciais dos três ajustamentos referentes aos graus 8, 10 e 12, não existem grandes diferenças, exceto na curva verde, correspondente ao grau 12, que está ligeiramente para cima. Nas pontas finais, a maior diferença está na curva azul, que corresponde ao ajustamento de grau 10, e que tem um valor mais baixo. Como todas as pontas iniciais têm valores mais baixos que as pontas finais, percebe-se que o ajustamento que terá um melhor comportamento é o ajustamento de grau 10, identificado pela curva azul.

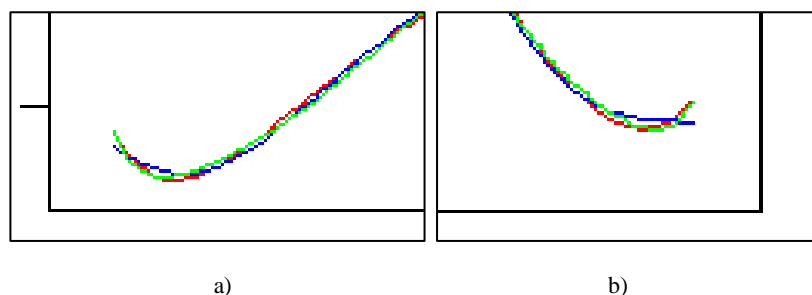


Figura 4.6 – a) Pontas iniciais das curvas dos ajustamentos polinomiais de grau 8, 10 e 12.
b) Pontas finais das curvas dos ajustamentos polinomiais de grau 8, 10 e 12.

Para além da análise que nos permite descobrir qual o melhor grau de ajustamento polinomial, vai ser feita uma análise ao conjunto de dados atual que inclui os anos desde 1977 até 2015. Como as temperaturas têm vindo a mudar lentamente ao longo dos anos, acredita-se que deixa de fazer sentido utilizar um conjunto tão alargado de anos, pelo que se vai apenas analisar o conjunto dos últimos dez anos e dos últimos vinte anos de temperaturas.

Assim, reuniram-se as respetivas médias e procedeu-se à mesma análise. Preferiu-se fazer primeiro a análise ao grau de ajustamento para se ter uma melhor perceção de qual o melhor grau e testar com os possíveis conjuntos de dados.

Tabela 4.4 – Significância dada pelo teste *t-student* aos coeficientes para o ajustamento polinomial de grau 12, para os diferentes conjuntos de dados dos últimos 10 anos e dos últimos 20 anos.

* - existe significância

N – não existe significância

Grau Polinómio	Significância dos Coeficientes							
	≤ 5	6	7	8	9	10	11	12
Últimos 10 anos	*	*	N	*	N	N	N	N
Últimos 20 anos	*	*	*	*	N	*	N	N

Na tabela 4.4, tem-se os ajustamentos com os novos conjuntos de dados, e pode observar-se que no primeiro conjunto todos os coeficientes com grau abaixo de 6 mostram significância e ainda o coeficiente de grau 8. Enquanto que no conjunto de dados correspondente aos últimos vinte anos, todos os coeficientes com grau abaixo do 8 e com grau 10 revelam significância.

Tal como nos casos anteriores, a diferença entre os coeficientes de determinação é muito pequena e todos se mantêm acima do valor 0.9. Comparando os resultados podemos concluir que o melhor grau para o conjunto dos últimos dez anos é o ajustamento de grau 8, pois o de grau 6 pode provocar perda de informação significativa e, no caso dos últimos vinte anos, o melhor ajustamento será de grau 8 ou 10. Esta informação confirma-se graficamente, nas seguintes figuras:

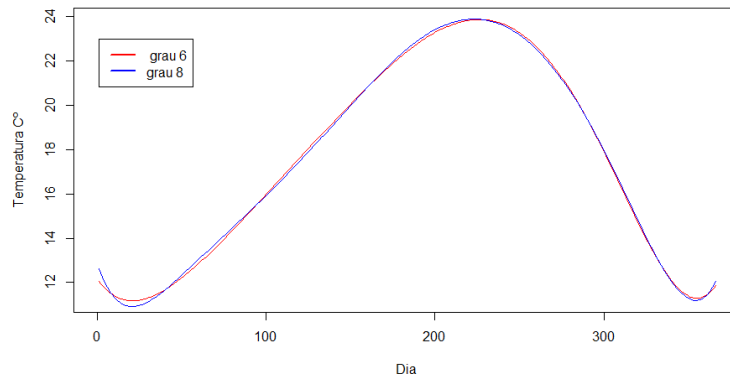


Figura 4.7 – Ajustamentos polinomiais de grau 6 e 8 para o conjunto de dados dos últimos 10 anos.

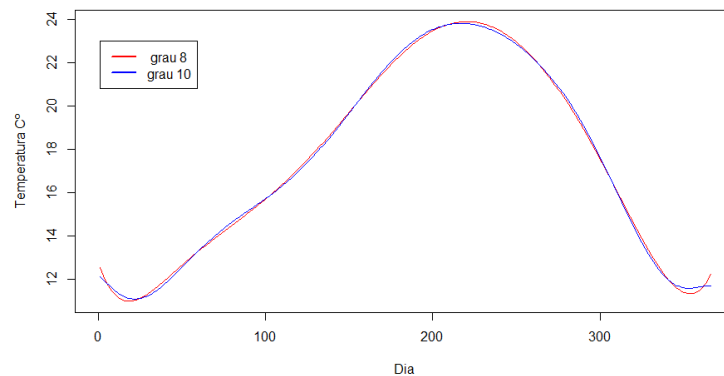


Figura 4.8 – Ajustamentos polinomiais de grau 8 e 10 para o conjunto de dados dos últimos 20 anos.

Observando a figura 4.7, confirma-se que a curva mais suave é a do ajustamento de grau 6, apresentando um melhor comportamento nas pontas do gráfico do que a curva de grau 8 mas, em contrapartida, há uma perda de informação, pois ao longo da curva perdem-se certas oscilações que explicam o comportamento dos dados. Já na figura 4.8, que corresponde aos dados dos últimos vinte anos, tem-se a presença das oscilações ao longo da curva nos dois ajustamentos, e nas pontas é possível observar-se que no ajustamento de grau 8 há uma grande diferença entre a ponta inicial e a ponta final. No ajustamento de grau 10 esta união é mais suave, pelo que este é o melhor grau para o ajustamento.

Comparando o conjunto de todos os anos com ajustamento de grau 10, o conjunto dos últimos dez anos com ajustamento de grau 8 e o conjunto dos últimos vinte anos com ajustamento de grau 10, tem-se as seguintes figuras:

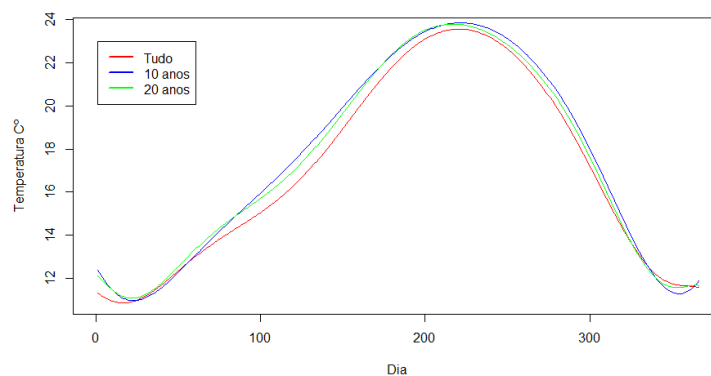


Figura 4.9 – Ajustamentos polinomiais do conjunto de todos os dados, conjunto últimos 10 anos e conjunto últimos 20 anos com os graus mais adequados respetivos.

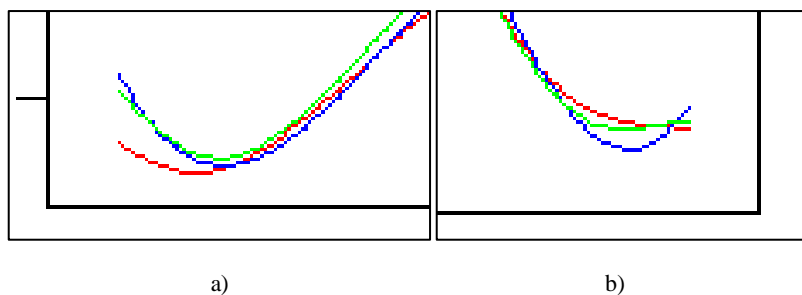


Figura 4.10 – a) Pontas iniciais dos ajustamentos polinomiais da figura 4.9.
b) Pontas finais dos ajustamentos polinomiais da figura 4.9.

Comparando os 3 conjuntos de dados com os respectivos graus de polinômio mais corretos, na figura 4.9, observa-se que existe uma diferença entre os conjuntos, pois vê-se que o conjunto de todos os dados apresenta valores mais baixos ao longo do ano. Isto mostra-nos que as temperaturas médias têm vindo a subir, pelo que será melhor optar por um dos novos conjuntos, de forma a manter os dados mais atualizados.

Analisando a figura 4.10, verifica-se que as pontas do conjunto dos últimos dez anos têm grandes oscilações, ao contrário do conjunto dos últimos vinte anos, pelo que a melhor decisão, tomada em conjunto com a empresa, será optar pelo conjunto de dados dos últimos 20 anos com um ajustamento de grau 10.

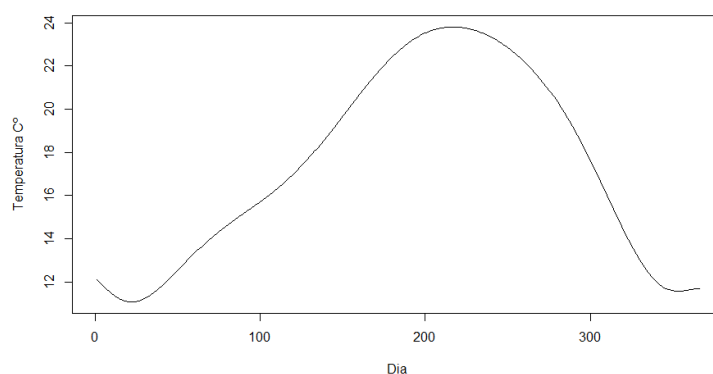


Figura 4.11 – Ajustamentos polinomial de grau 10 para o conjunto dos últimos 20 anos.

Como ainda existem algumas falhas na junção das pontas (a linha a preto na figura 4.12 é a curva da figura 4.11 no intervalo de tempo de 18 de julho a 17 de julho), decidiu-se fazer um ajustamento polinomial aos mesmos dados com grau muito elevado, mas nesta nova ordem cronológica, para corrigir a mudança abrupta de temperatura e não perder informação fundamental.

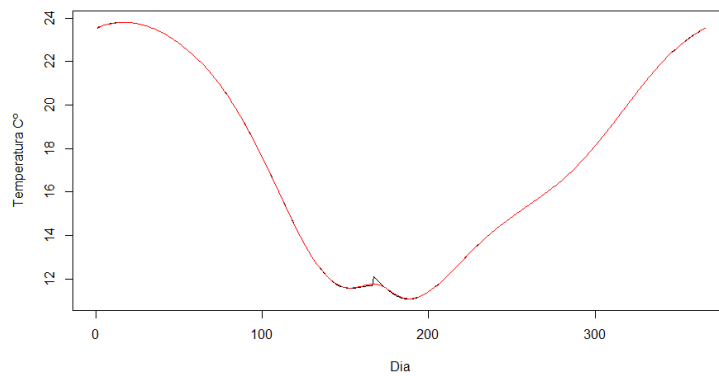


Figura 4.12 – Ajustamento à junção das pontas do anterior ajustamento polinomial.

Desta forma, tem-se na figura 4.13 a curva para a nova temperatura normal, obtida das médias das temperaturas dos últimos anos com um ajustamento de grau 10.

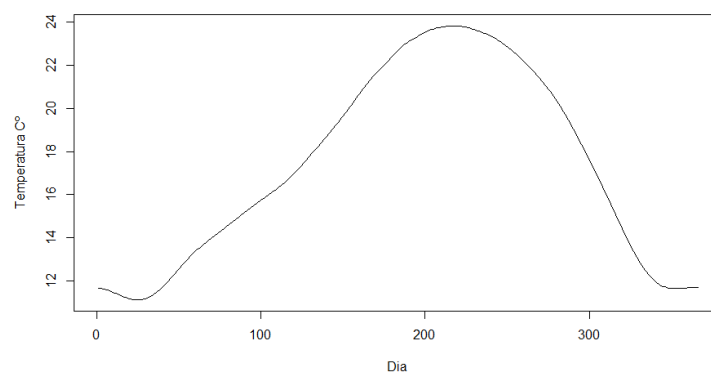


Figura 4.13 – Temperaturas normais finais.

b. Gradientes

Nesta fase do projeto pretende-se melhorar o cálculo dos gradientes, através da reavaliação do processo da sua obtenção, ajustando-o e aperfeiçoando-o de modo a continuar a ter um erro baixo, sem deixar de respeitar o comportamento dos gradientes ao longo do ano.

Os gradientes são o declive da reta de regressão entre os desvios de consumo e as previsões dadas pelo modelo e os desvios de temperatura para a temperatura normal e, tal como referido anteriormente, conta-se com três gradientes distintos que são divididos consoante o desvio de temperatura. Neste processo será utilizado todo o conjunto de dados disponibilizado, desde o ano de 2002 até 2015.

Tem-se como objetivo manter a relação entre os dois tipos de desvios como base dos gradientes, mas num processo iterativo encontrar a convergência dos gradientes, ou seja, em cada iteração calculam-se novos gradientes com base nas previsões geradas pelos gradientes da iteração anterior e assim por diante, até se obter convergência, verificada graficamente. Numa abordagem mais teórica, planeia-se construir uma função que calcule os três gradientes, através dos desvios do consumo e dos desvios de temperatura, para cada semana. Com esses novos gradientes, calculam-se as novas previsões que vão resultar em novos desvios de consumo e, conseqüentemente, novos gradientes. Durante este processo, é sempre calculada e comparada a medida de desempenho RMSE. Não é utilizado o erro aplicado pela empresa, o erro percentual de desvio anual, pois trata-se de um desvio contabilizado por ano, e neste caso, pretende-se comparar uma única medida para todo o conjunto de dados.

Para atingir este objetivo foram criadas três funções para todo o processo. A função principal, *funmoddif*, que engloba as funções *gradientes* e *tratamento* (anexos, pp. 57). Na primeira função é realizado o processo iterativo descrito anteriormente, tendo como gradientes iniciais os utilizados atualmente pela empresa. Dentro do ciclo iterativo são calculados os declives da relação dos desvios, na função *gradientes*, de seguida suavizados com um ajustamento polinomial, de forma a encontrar o erro menor de previsão (RMSE) dentro dos possíveis ajustamentos.

Na função *gradientes*, caso não existam pontos ou exista apenas um ponto na seleção consoante o desvio de temperatura, o valor do gradiente fica a zero, pois não é possível encontrar uma relação linear com esse número de pontos.

Posteriormente é feita uma verificação gráfica da convergência de cada gradiente. Caso esta não seja verificada, o processo repete-se até ser encontrada a convergência nos três gradientes. A função *tratamento* é uma função usada inicialmente para validar os dados e juntá-los em apenas um *data.frame* de forma a simplificar o seu uso no código.

Começou-se por fazer testes para descobrir o melhor intervalo de graus para o ajustamento polinomial.

Tabela 4.5 – RMSE de previsão e graus obtidos para o ajustamento polinomial dos três gradientes, para cada intervalo testado.

Intervalo	0 a 5	5 a 10	10 a 15
RMSE	0.04858	0.04875	0.04879
Grau	[2 4 3]	[5 6 5]	[13 11 12]

Observando os resultados deste teste inicial na tabela 4.5, tem-se que o RMSE apenas difere na quarta e quinta casa decimal, tendo o intervalo [0,5] o RMSE mais baixo. Apesar da pequena diferença entre os erros, o melhor intervalo de graus para o ajustamento é o mais pequeno, 0 a 5. Uma avaliação visual do gráfico dos gradientes revela que este intervalo é o que apresenta mais suavidade enquanto os restantes apresentam muitas oscilações não desejadas, que para além disso contribuem para um erro maior.

Para cada gradiente tem-se os seguintes graus:

- Gradiente 1 – 0, 2, 3, 4, 5;
- Gradiente 2 – 2, 3, 4, 5;
- Gradiente 3 – 2, 3, 4, 5.

No gradiente 1, inclui-se o grau 0 que corresponde a utilizar o conjunto de valores do gradiente atual. Neste e nos restantes não se utiliza o grau 1, pois um ajustamento polinomial deste grau corresponde a uma reta, que não se adapta ao comportamento observado. Ao longo do ano o gradiente deve seguir no mínimo o comportamento de uma parábola e os valores iniciais e finais devem ter o mesmo sinal (positivo ou negativo).

De seguida foram feitos testes para comparar o RMSE para as diferentes combinações de temperaturas normais e gradientes anteriormente obtidos.

Tabela 4.6 – RMSE de previsão para cada combinação de temperaturas normais e gradientes.

	RMSE
Velhos Gradientes e Velhas Temperaturas	0.049708
Novos Gradientes e Velhas Temperaturas	0.049672
Velhos Gradientes e Novas Temperaturas	0.049099
Novos Gradientes e Novas Temperaturas	0.048577

Mais uma vez, os resultados dos RMSEs variam muito pouco, mas é notável ver que para os novos gradientes com as novas temperaturas normais verifica-se uma melhoria. Além desta melhoria, também se observa que os resultados relativos às novas temperaturas normais são melhores independentemente dos gradientes utilizados, o que mostra que era necessária uma atualização a estas temperaturas.

Nas figuras 4.14, 4.15 e 4.16, apresentam-se os gráficos que mostram a convergência de cada gradiente, estando a preto o gradiente atual (“velho”). Como se consegue observar, o contradomínio dos diversos gráficos é muito semelhante, mas temporalmente existem algumas diferenças principalmente no gradiente 1. No entanto, conseguem ver-se alguns ajustes, pois tornou-se relevante ter novas temperaturas que se foram alterando durante o tempo, e como consequência, o consumo também se alterou. Além destas mudanças, graficamente atingiu-se o visual pretendido para os gradientes, sem que estes apresentassem oscilações abruptas.

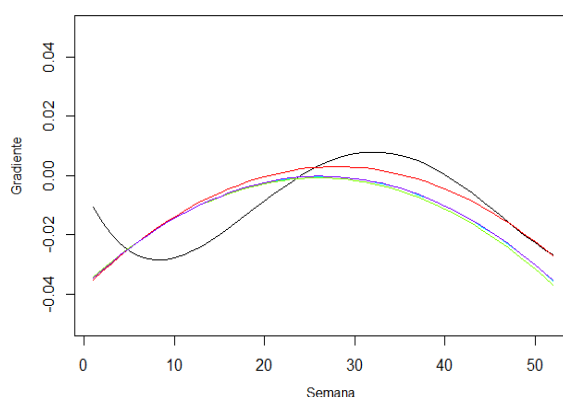


Figura 4.14 – Evolução e convergência do gradiente 1.

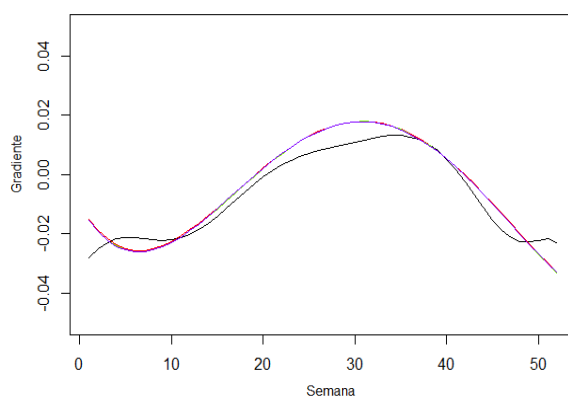


Figura 4.15 – Evolução e convergência do gradiente 2.

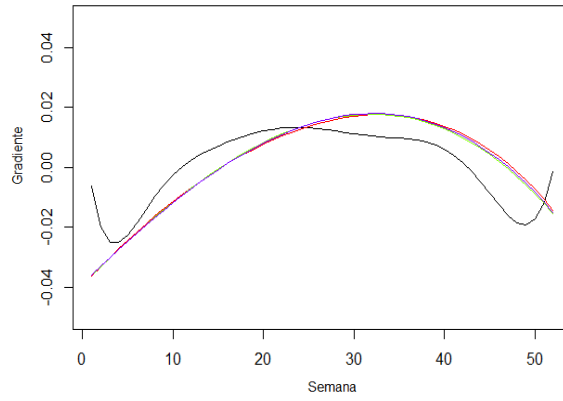


Figura 4.16 – Evolução e convergência do gradiente 3.

Nas figuras 4.17 e 4.18 estão representados o fator de correção da semana um e da semana cinquenta e dois, respetivamente, que são obtidos através dos gradientes (ver fórmula 2.23). Através da interpretação dos eixos dos gráficos tem-se que se G_{ikj} for 0 então o fator de correção é 1, isto é, vendo para o exemplo do gradiente 2:

$$\text{Fator de correção: } \frac{1}{1 + G_{ikj}} = \frac{1}{1 + 0} = 1 \quad (4.2)$$

$$\begin{aligned} G_{ikj} &= g_{1k} * \text{Min}(0, \phi_{ikj} - g_{\text{inf}}) + \\ &g_{2k} * \text{Max}(\text{Min}(\phi_{ikj}, g_{\text{sup}}), g_{\text{inf}}) + \\ &g_{3k} * \text{Max}(0, \phi_{ikj} - g_{\text{sup}}) = 0 \\ \Leftrightarrow g_{2k} * \text{Max}(\text{Min}(\phi_{ikj}, g_{\text{sup}}), g_{\text{inf}}) + 0 + 0 &= 0 \\ \Leftrightarrow g_{2k} * \text{Max}(\text{Min}(\phi_{ikj}, g_{\text{sup}}), g_{\text{inf}}) &= 0 \\ \Leftrightarrow \phi_{ikj} &= 0 \end{aligned} \quad (4.3)$$

Se G_{ikj} for 0, significa que o desvio de temperatura é 0, pois $g_{\text{sup}} = 2$ e $g_{\text{inf}} = -2$. Assim, no gráfico, o eixo das abcissas corresponde aos desvios da temperatura e no eixo das ordenadas, tem-se um fator de correção associado a esse desvio.

O importante dos dois gráficos é verificar que existe alguma similaridade no comportamento dos três fatores, onde o gradiente i ($i=1,2,3$) tem o mesmo sinal para a semana um e para a semana cinquenta e dois, que correspondem à primeira semana de janeiro e última semana de dezembro. No caso da semana um, o fator de correção no gradiente 1 e no gradiente 3 é mais forte que no gradiente 2, pois têm uma maior inclinação, logo o gradiente é menor. Na semana cinquenta e dois verifica-se o inverso para o gradiente 3, o fator de correção é mais fraco quando comparado com o gradiente 2.

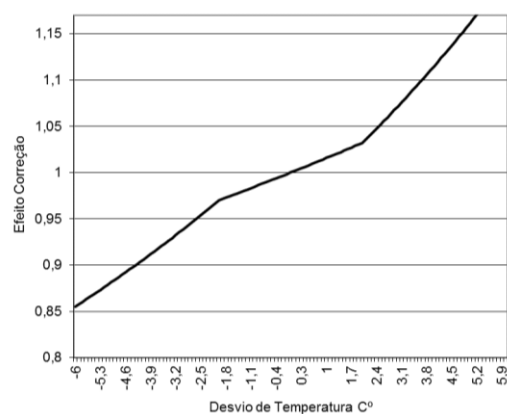


Figura 4.17 – Efeito fator de correção da semana 1.

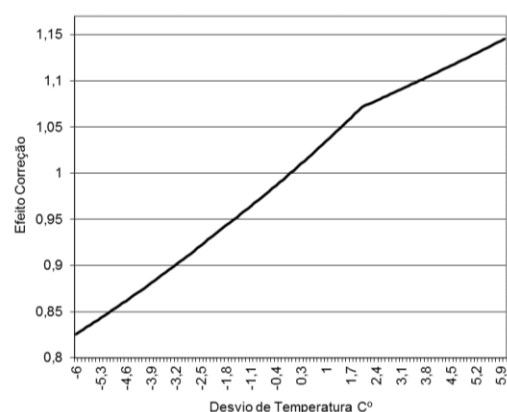


Figura 4.18 – Efeito fator de correção da semana 52.

Na seguinte tabela estão os erros de previsão anuais em percentagem do valor real (fórmula 2.19), com os novos gradientes. Apesar de ao longo dos anos se encontrarem erros um pouco altos em valor absoluto relativamente aquilo que se pretende, o ano a avaliar é o de 2015, que é o mais recente. Este apresenta um erro bastante baixo, o que se poderá dever a contaminação, pois os dados de 2015 também foram utilizados na calibração do modelo. Para além de 2015, os erros dos anos iniciais também apresentam valores baixos, sendo apenas os anos intermédios os que exibem valores mais altos em valor absoluto.

Tabela 4.7 – Erro anual em percentagem de previsão de 2003 até 2015.

Erro anual em % do valor previsto com valores históricos até 2015												
2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
-1.26	-0.37	1.03	-1.55	-1.66	-2.42	-4.24	4.48	-3.63	-4.39	3.84	2.76	0.25

Um último teste foi feito com os dados históricos desde 2002 até 2014, ou seja, não se incluiu os dados de 2015, para que as previsões não tivessem qualquer relação com os dados de 2015.

Tabela 4.8 – RMSE de previsão e erro anual em percentagem de previsão sem valores históricos de 2015.

	RMSE	Erro % 2015
Erro com Novas Temperaturas	0.04850	0.4363

Em comparação com o teste anterior, o erro percentual de 2015 é mais alto, como era de esperar; no entanto, o valor do erro continua baixo, revelando-se abaixo de 1%. A diferença entre os RMSEs é muito pequena.

Após a análise da convergência dos gradientes, consegue-se perceber que existem pequenas melhorias, o que confirma a necessidade e importância de ajustar e corrigir os valores em estudo, garantindo uma melhor previsão do consumo, tendo como base, a evolução das temperaturas climatéricas ao longo dos anos.

II. Luminosidade

Neste subcapítulo pretende-se estudar a hipótese de inclusão de uma medida de luminosidade no modelo. Para tal é necessário perceber se existe alguma relação entre a luminosidade e o consumo de eletricidade, ao ponto de influenciar o consumo e justificar a inclusão desta medida. Neste sentido, começa-se por uma análise às séries de consumo de eletricidade e de luminosidade e, de seguida, procede-se a uma análise de correlação.

Os painéis fotovoltaicos captam a luminosidade e produzem energia elétrica. Pode-se considerar então que a luminosidade absoluta é diretamente proporcional à energia produzida. Para a medir em termos relativos, utiliza-se o quociente entre a energia produzida e a capacidade total da potência instalada.

Os dados do consumo de eletricidade são uma série de consumo diário corrigido da temperatura, e os dados da luminosidade são os relacionados com potências instaladas nas centrais de produção fotovoltaica em território nacional, em valor absoluto, e em percentagem da capacidade de cada central (potência instalada). Ao longo dos anos, estas centrais têm vindo a aumentar, quer em número quer em capacidade, pelo que o mais correto é trabalhar em termos relativos, ou seja, com a potência em percentagem da capacidade de cada central.

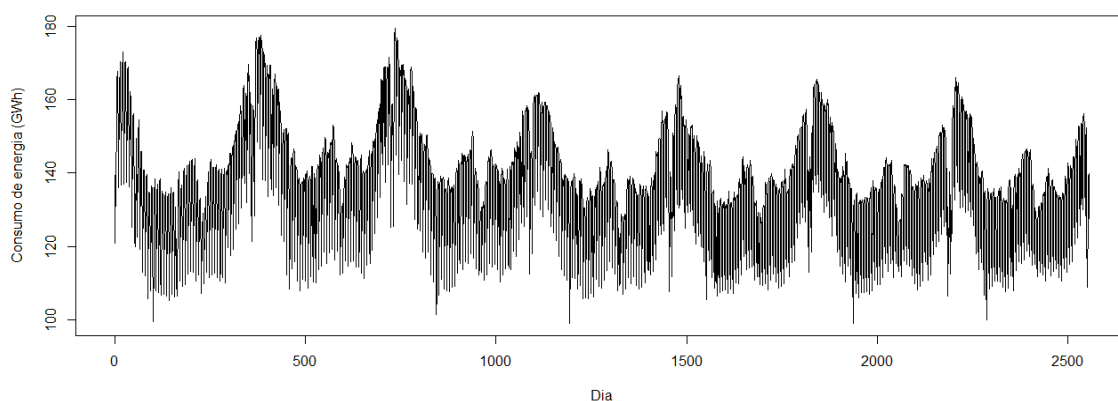


Figura 4.19 – Série de consumo de eletricidade de 2009 até 2015.

Na figura 4.19, apresenta-se a série do consumo de eletricidade, com dados a nível nacional desde o ano de 2009 até 2015. Analisando o gráfico, consegue-se observar uma ligeira tendência negativa, o que significa que o consumo tem vindo a descer ao longo dos anos. Além desta tendência, são visíveis vários períodos de sazonalidade, confirmados no periodograma da figura 4.20, sendo os períodos mais significativos de 365 dias (um ano) e de 7 dias (semanal).

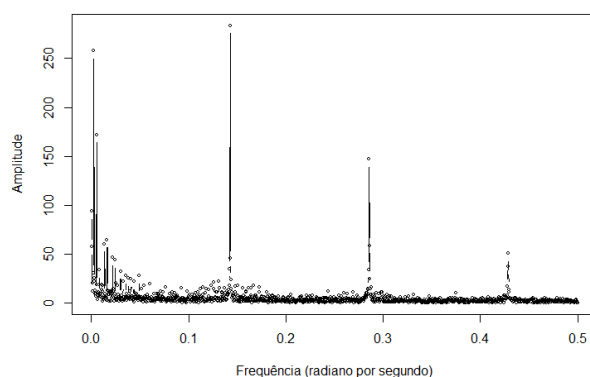


Figura 4.20 – Periodograma da série de consumo de eletricidade apresentado na figura 4.19.

Nas figuras seguintes, figuras 4.21 e 4.22, tem-se as séries correspondentes à luminosidade utilizada para a produção de eletricidade, a primeira em valor absoluto e a segunda em medida relativa.

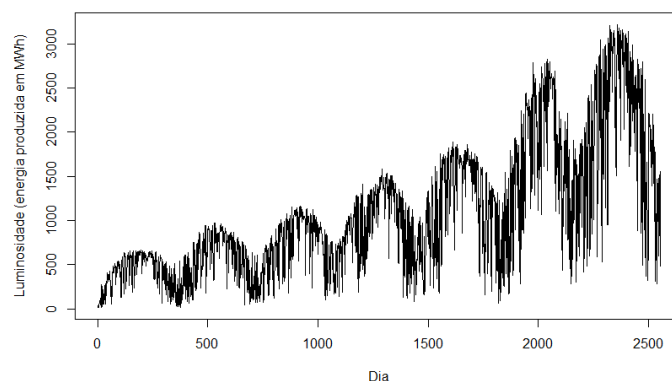


Figura 4.21 – Série de luminosidade registrada de 2009 até 2015.

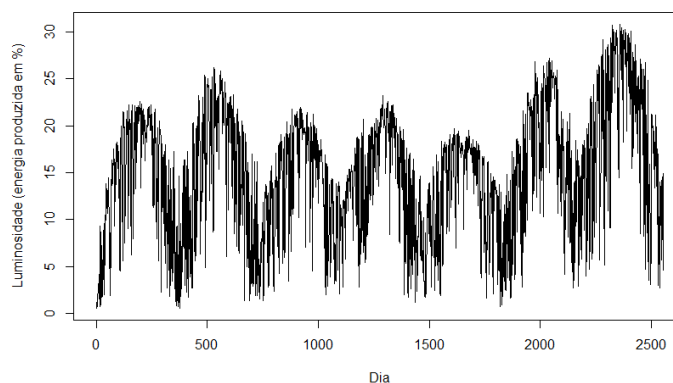


Figura 4.22 – Série de percentagem de luminosidade registrada de 2009 até 2015.

Analisando estas figuras, observa-se que no primeiro caso, existe uma clara tendência, consequência da potência produzida ter vindo a aumentar “exponencialmente”, devido ao aumento da capacidade da potência instalada; no segundo caso, nota-se que a tendência começa por ser decrescente, e nos últimos dois anos inverte-se, aumentando significativamente, chegando a ultrapassar os 30% de capacidade total. Além desta tendência crescente, pode-se verificar que existe, pelo menos, um período de sazonalidade, que corresponde ao período de 365 dias, que se confirma no periodograma.

Para uma análise mais detalhada relativamente ao comportamento da série durante um ano, apresenta-se o gráfico do ano 2015 na figura 4.23, que mostra tendência ao longo do ano, atingindo o seu máximo no período do solstício de verão, altura do ano em que o sol é mais forte e existe menos nebulosidade na atmosfera. O periodograma desta série mostra a existência de dois períodos pouco significativos correspondentes a 20 dias e a 60 dias.

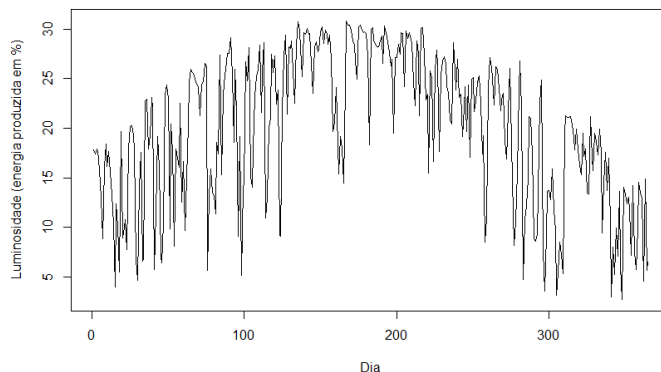


Figura 4.23 – Série de percentagem de luminosidade registada no ano 2015.

Visto que existe pelo menos um período de sazonalidade em comum entre a série de consumo e a série de percentagem de potência em que ambos exibem um dado comportamento no mesmo período de tempo, pode-se suspeitar da existência de alguma relação entre os dois conjuntos de dados. Para analisar esta hipótese, procedeu-se a uma análise de correlação.

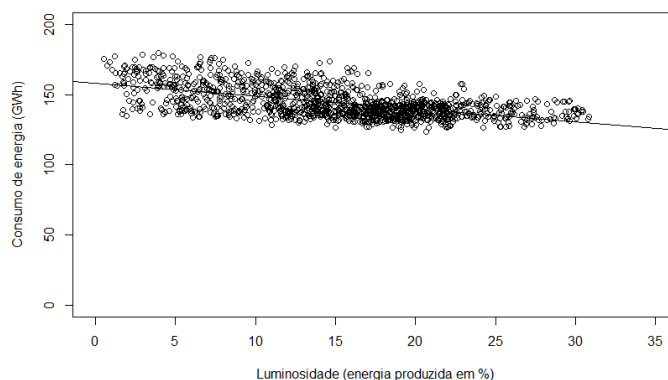


Figura 4.24 – Diagrama de dispersão entre consumo de eletricidade e percentagem de luminosidade em dias úteis.

Na figura 4.24, observa-se o diagrama de dispersão entre o consumo e a luminosidade em dias úteis, pois o comportamento do consumo difere entre os vários dias da semana. Pela nuvem de pontos, consegue-se ver uma certa inclinação dos mesmos, isto é, quanto menor a luminosidade, mais alto o consumo, e vice-versa. Mesmo que esta relação não seja evidente na nuvem de pontos, confirma-se esta relação com o declive da reta da regressão.

Na tabela 4.9, mostra-se o coeficiente de correlação para vários casos. Na primeira linha apresenta-se a correlação considerando todo o conjunto de dados disponíveis, enquanto na segunda linha considera-se apenas os dados dos meses de janeiro e fevereiro (inverno). A potência instalada para calcular a percentagem é apenas registada no primeiro dia de cada ano (1 de janeiro) pelo que não é “correto” utilizar este número para calcular a luminosidade relativa dos meses mais afastados. No entanto foi elaborado o gráfico da figura 4.24, para se ter uma noção geral do comportamento do ano inteiro e não apenas de um mês.

Tabela 4.9 – Coeficiente de correlação entre consumo de eletricidade e percentagem de luminosidade, dividido por tipo de dia.

Dias Úteis	Sábados	Domingos
-0.56187	-0.49230	-0.60275
Uteis Inv.	Sáb. Inv.	Dom.Inv.
-0.49764	-0.35218	-0.60941

Analisando os coeficientes de correlação, verifica-se imediatamente que os coeficientes da primeira linha são mais altos em valor absoluto do que os da segunda linha, com uma exceção: os domingos de inverno, em que os coeficientes de correlação são muito semelhantes nos dois grupos.

No geral, os coeficientes são moderados e de valor negativo, mas não despreciables, talvez com a exceção dos sábados de inverno.

De forma a ter uma análise mais completa, com todos os meses, decidiu-se obter as percentagens através de outro método: obteve-se a capacidade de algumas potências instaladas e espalhadas por Portugal, registou-se a produção diária através da luminosidade para cada central e calculou-se a percentagem diária com a respetiva capacidade de potência instalada. Por fim, calculou-se a média diária das percentagens de todas as potências. Com esta metodologia, passa-se a ter um maior equilíbrio entre os dados, pois existem potências com capacidades muito diferentes.

As séries relativas à energia fotovoltaica produzida através de luminosidade são constituídas por dados de dez potências do Alentejo, uma de Trás-os-Montes, uma de Lisboa, uma do Ribatejo e uma do Algarve. De seguida, analisou-se a correlação apenas para algumas das potências instaladas: quatro grandes do Alentejo, uma de Trás-os-Montes, uma de Lisboa e uma do Ribatejo. Outra análise foi feita só aos dados da potência localizada em Lisboa. Por fim, uma análise ponderada, apenas com uma potência do Alentejo, uma de Trás-os-Montes, uma de Lisboa e uma do Ribatejo, onde os dados de Trás-os-Montes e do Ribatejo são multiplicados por 2, e os de Lisboa por 3 (análise baseada na densidade populacional).

As análises anteriores serviram para verificar a influência da localização das potências instaladas na relação entre a percentagem de luminosidade e o consumo, experimentando todas ou algumas potências, ou uma média ponderada, com o intuito de estudar qual é o melhor conjunto de dados para a construção da medida.

Tabela 4.10 – Coeficiente de correlação para cada análise, dividida por tipo de dia.

Coef. Correlação	Total	Dias úteis	Sábados e Domingos
Total	-0.36049	-0.54844	-0.52798
Algumas	-0.35969	-0.54431	-0.52648
Lisboa	-0.29354	-0.44625	-0.43144
Ponderada	-0.34828	-0.52267	-0.50939

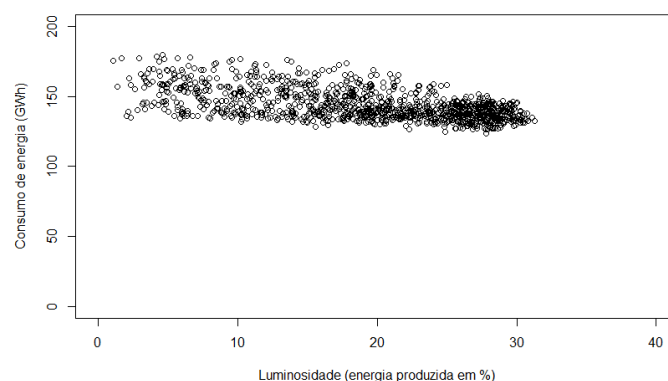


Figura 4.25 –Diagrama de dispersão entre o consumo de eletricidade e percentagem de luminosidade de todas as centrais em dias úteis.

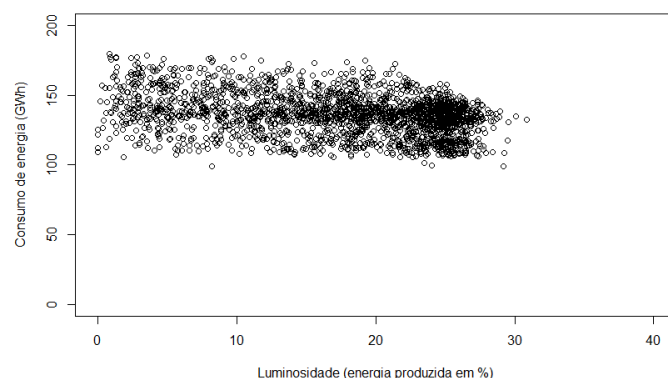


Figura 4.26 – Diagrama de dispersão entre o consumo de eletricidade e percentagem de luminosidade da central em Lisboa.

Verifica-se que o caso de Lisboa é bastante diferente dos restantes: os valores da correlação baixam em valor absoluto de forma significativa e o diagrama de dispersão (ver figura 4.26) apresenta uma maior entropia. Nos outros casos os valores são muito similares, independentemente do tipo de dia em estudo.

Os dados referentes aos dias úteis são os que têm valores mais baixos, seguindo-se a pequena distância os dados dos fins-de-semana. No entanto o valor global para todos os dados é bastante superior a estes, o que mostra que é importante a diferenciação dos dias na análise do consumo.

Destas últimas quatro análises, conclui-se que a que tem coeficientes de correlação mais significativos é a que inclui todas as potências instaladas escolhidas, no entanto estes valores são ainda mais altos do que os obtidos inicialmente, considerando todas as potências instaladas no país, pelo que foi este o conjunto de dados escolhido.

Tendo agora provado uma relação entre a luminosidade e o consumo de eletricidade, pretende-se determinar a luminosidade normal para cada dia para que a partir desta se possa efetuar um processo de correção no consumo. Para a determinação da luminosidade normal, o procedimento será similar ao da temperatura, que é feito através de um ajustamento polinomial à série das médias diárias das percentagens de utilização das potências fotovoltaicas instaladas.

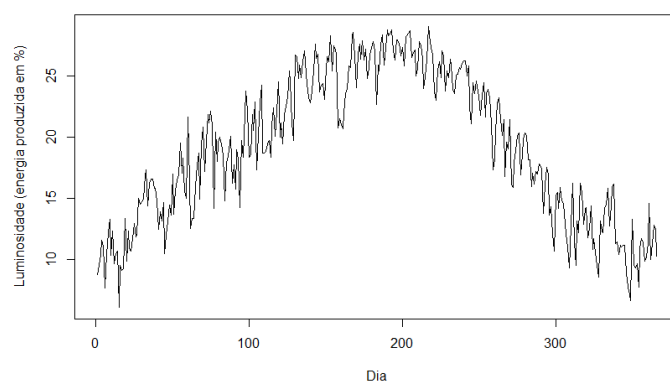


Figura 4.27 – Percentagem de luminosidade média diária anual.

Analisando a série de luminosidade média anual da figura 4.27, verifica-se que na altura do verão ela é máxima e apresenta muitas oscilações ao longo do ano, tal como a da temperatura média. Estas irregularidades devem-se principalmente ao facto de as centrais não estarem ligadas todos os dias e à existência de quebras. No entanto, não existem períodos de sazonalidade, de acordo com o periodograma, existindo apenas tendência.

Começou-se por fazer ajustamentos de grau 10 e grau 15, para ver quais os coeficientes que podem ser significativos para o modelo de regressão.

Tabela 4.11 - Significância dada pelo teste t-student aos coeficientes para cada ajustamento polinomial de grau 10 e 15.

* - existe significância

N – não existe significância

Grau Polinómio	Significância dos Coeficientes									
	≤2	3	4	5	6	7	8	9	10	≥11
Grau 10	*	N	*	*	*	*	N	N	*	-
Grau 15	*	N	*	*	*	*	N	N	*	N

Tabela 4.12 – Coeficiente de determinação (e ajustados) para cada ajustamento polinomial de grau 10 e 15.

	Grau 10	Grau 15
R^2	0.8921	0.8938
R^2 ajustados	0.889	0.8893

Mais uma vez, os coeficientes de determinação pouco diferem, cuja diferença apenas é visível na terceira casa decimal, tal como acontece com os coeficientes de determinação ajustados. Ainda assim nos coeficientes, consegue-se perceber que os mais significativos são os das potências com grau menor ou igual a 2 e com grau entre 4 e 7. O ajustamento com um polinómio de grau 2 é fraco e perder-se-ia informação. Acima do grau 8, apenas o coeficiente de grau 10 é significativo, tornando o modelo demasiadamente complicado.

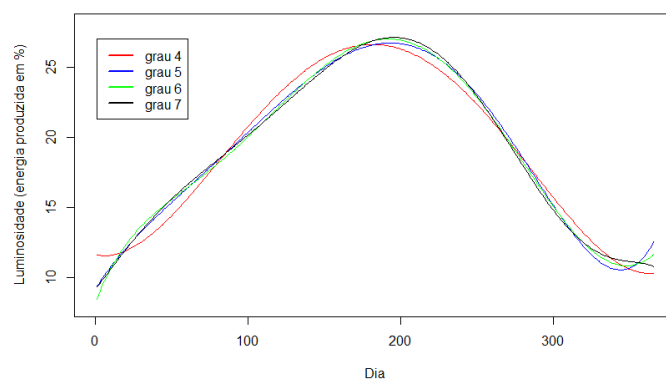


Figura 4.28 – Curvas dos ajustamentos polinomiais de grau 4, 5, 6 e 7.

Na figura 4.28, estão representadas as curvas dos quatro ajustamentos referidos anteriormente. A curva vermelha, correspondente ao grau 4, é eliminada, pois é notório no gráfico o desfasamento da mesma relativamente às restantes curvas. A curva azul (grau 5), também não revela um bom ajustamento, porque ao analisar as extremidades do gráfico, notamos que na ponta final há uma grande subida, ao contrário da ponta inicial, o que criará uma oscilação abrupta não desejável.

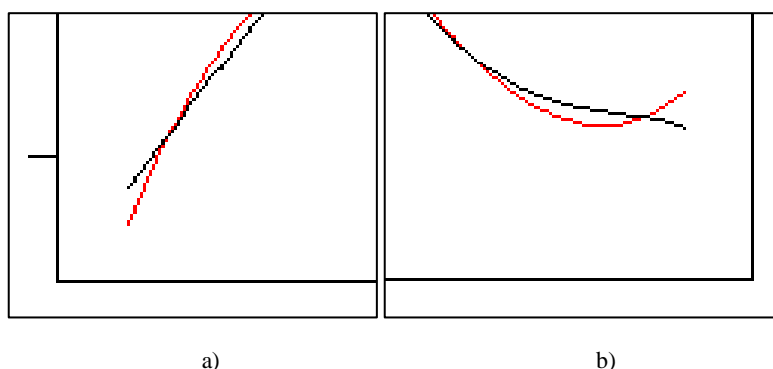


Figura 4.29 – a) Pontas iniciais dos ajustamentos polinomiais de grau 6 e 7.
b) Pontas finais dos ajustamentos polinomiais de grau 6 e 7.

Entre estes dois últimos casos, grau 6 e grau 7, o melhor é optar por este último, pois é aquele que mostra menos diferença entre a ponta inicial e final. Mais uma vez, é feito de novo um ajustamento polinomial aos dados com uma ordem cronológica diferente (intervalo de tempo de 18 de julho a 17 de julho) de forma a suavizar a junção das pontas e eliminar a oscilação indesejada, tal como foi feito para as temperaturas normais.

Como o comportamento da luminosidade se assemelha ao da temperatura, e determinados os valores de luminosidade normal representados na figura 4.30, implementar-se-á um método de correção ao consumo com base na luminosidade muito semelhante ao que foi efetuado à temperatura.

Dependendo dos resultados, irá procurar-se uma forma de melhorar a respetiva correção. No entanto, por limitações de tempo e condições alheias a esta análise, não foi possível continuar esta fase do projeto, pelo que será para um trabalho a ser executado no futuro. No entanto, no capítulo 5, correspondente à conclusão, serão dados mais pormenores permitindo uma melhor compreensão deste tema.

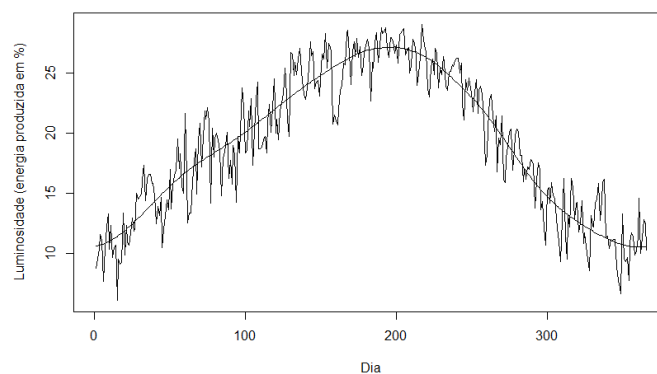


Figura 4.30 – Ajustamento polinomial de grau 7 e médias diárias de percentagem de luminosidade.

III. Gás

a. Relação

Com o objetivo de adaptar o modelo de previsão existente ao consumo de gás é necessário verificar se existe uma relação entre este e a temperatura e os tipos de dias, de forma a verificar se o modelo utilizado pela empresa pode ser adaptado. No entanto, ao contrário do consumo de eletricidade, os dados históricos do consumo de gás natural de mercado convencional sem alta pressão começam a partir de 2009, e com absoluta credibilidade apenas depois de 2012/2013, pelo que o histórico será bastante pequeno neste subcapítulo.

Tendo este objetivo em mente, começa-se por testar a série do consumo de gás a nível de sazonalidade, tendência e outros componentes da análise de séries temporais.

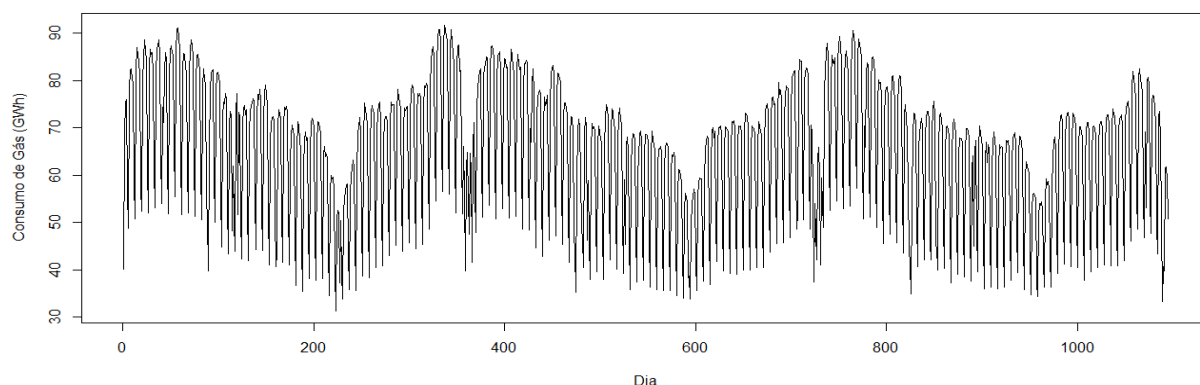


Figura 4.31 – Série de consumo de gás natural de 2013 até 2015.

Observando a figura 4.31, verifica-se apenas a existência de sazonalidade com períodos anual e semanal. O periodograma da figura 4.32 confirma estes dois períodos de sazonalidade, correspondendo aos dois picos mais altos da figura indicada, sendo o mais forte o período de 7 dias e o seguinte o período de 365 dias. Além destes períodos existem outros com significância muito inferior, daí não serem visíveis graficamente na figura 4.31. O periodograma não mostra a existência de tendência que, usualmente, é assinalada pelo pico mais alto corresponder ao valor 0 de frequência. (Note-se que o pico do período de sazonalidade anual não é exatamente na frequência nula.)

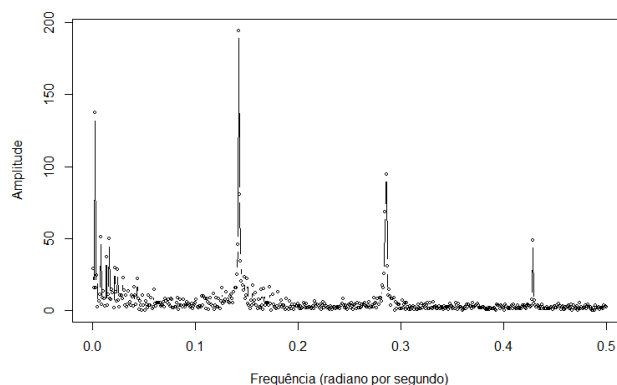


Figura 4.32 – Periodograma da série de consumo de gás.

Ao analisar a figura seguinte, que representa o periodograma da série da temperatura no mesmo intervalo de tempo, verifica-se também o mesmo período de sazonalidade anual e ausência de tendência.

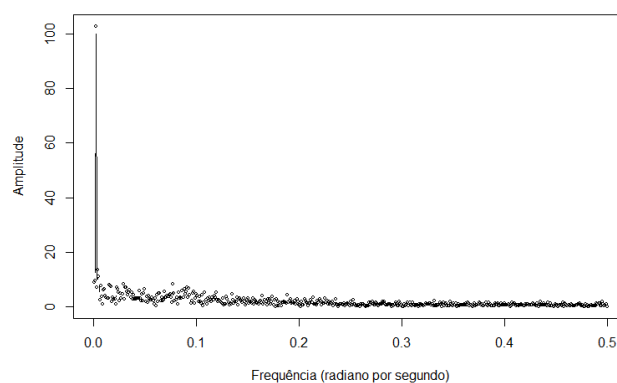


Figura 4.33 – Periodograma da série de temperatura registrada.

Dado existir um período de sazonalidade comum entre as duas séries, pode-se pôr a hipótese de existência de uma relação entre essas duas variáveis. As características da série de consumo de gás natural, com a presença de sazonalidade e ausência de tendência, sugerem-nos a aplicação do mesmo modelo de previsão usado para a eletricidade, o modelo de Holt-Winters.

No entanto, é necessário testar o tipo de relação que existe entre a temperatura e o consumo de gás, e a diferenciação do consumo pelas diversas classes de tipos de dia, a que chamamos efeito de calendário. Existem dez classes de tipo de dia, entre os quais estão incluídos os dias úteis, sábados e domingos, feriados, “pontes” e outros dias que possam influenciar o consumo-padrão. Para este caso vai-se estudar inicialmente os dias úteis (segundas-feiras à parte), sábados, domingos e feriados. Após esta primeira fase de estudo pretende-se refletir se faz sentido avançar com o estudo para as restantes classes de tipo de dia, visto que o número de dias incluídos nessas é muito reduzido para uma análise conclusiva, ou compreender se se elimina alguma delas.

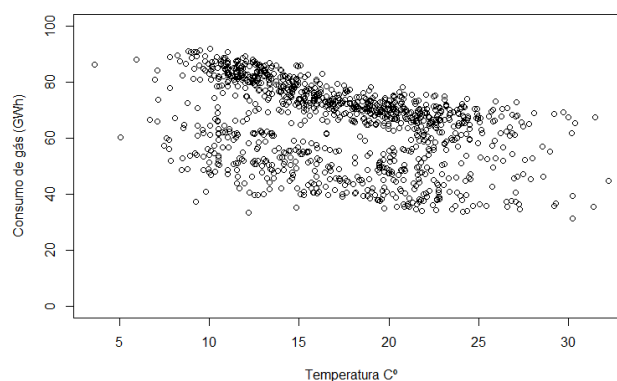


Figura 4.34 – Diagrama de dispersão entre consumo de gás e temperatura.

Na figura 4.34, apresenta-se o diagrama de dispersão relativamente aos dados da temperatura e do consumo do gás. Observa-se uma ligeira associação inversa entre as variáveis, que se confirma com o valor do coeficiente de correlação de Pearson, na tabela 4.13. Analisando os coeficientes, ano a ano, observam-se valores similares ao valor geral.

Tabela 4.13 – Coeficiente de correlação total e para cada ano.

ρ Total	ρ 2013	ρ 2014	ρ 2015
-0.3887	-0.4107	-0.3577	-0.4058

Tabela 4.14 – Coeficiente de correlação para cada ano diferenciado o conjunto de temperaturas normais.

Temp. Normal	ρ 2013	ρ 2014	ρ 2015
Antiga	-0.4252	-0.4061	-0.3672
Nova	-0.4308	-0.4118	-0.3749

Na tabela 4.14 apresentam-se os coeficientes de correlação do consumo de gás relativamente aos dois conjuntos de temperaturas normais: as utilizadas atualmente pela empresa e as obtidas através do ajustamento polinomial. Como se pode ver os valores são muito semelhantes, existindo uma melhoria para as novas temperaturas normais.

Na figura 4.35, analisa-se a relação consumo/temperatura discriminando os dias pela estação do ano a que pertencem, cada uma com uma cor distinta. Os pontos de cor preta e azul correspondem ao inverno e ao verão, respetivamente, e mostram uma separação bastante distinta entre os dois grupos. Não se nota, graficamente, uma relação entre as variáveis dentro destas estações, pois os pontos apresentam um comportamento algo vertical (nuvem aproximadamente esférica/oval). As cores verde e vermelho correspondem à primavera e ao outono, onde se verifica uma maior dispersão em relação às outras duas estações do ano, sendo possível notar a existência de alguma correlação negativa, pois a nuvem desenvolve-se na diagonal, de NW para SE. No geral verifica-se que os consumos associados ao verão são mais baixos (sendo as temperaturas mais altas), verificando-se comportamento contrário no inverno.

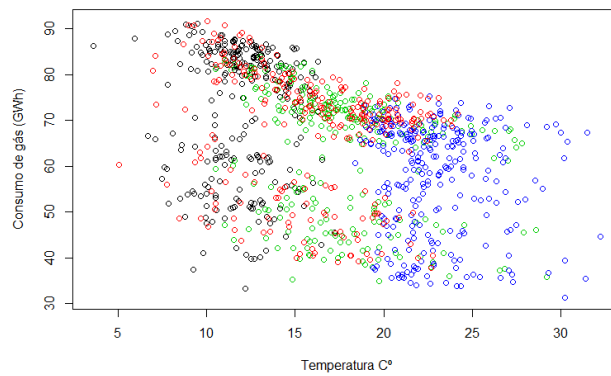


Figura 4.35 – Diagrama de dispersão entre o consumo de gás e as temperaturas, dividido pelas estações do ano.

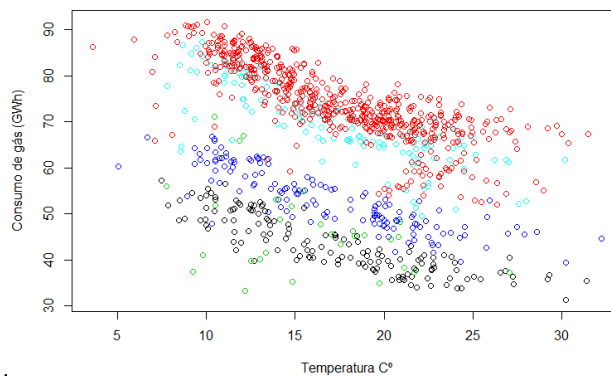


Figura 4.36 – Diagrama de dispersão entre o consumo de gás e as temperaturas, dividido pelo tipo de dia.

Na figura 4.36 apresenta-se o consumo referente aos cinco tipos de dia citados anteriormente. Note-se que os valores mais altos, a vermelho e azul-bebé, correspondem aos dias úteis e à segunda-feira, respetivamente, enquanto as restantes cores, preto, azul e verde, correspondem aos sábados, domingos e feriados.

O gráfico seguinte apresenta a média de consumo de gás para cada tipo de dia, no universo dos três anos identificados. Como se consegue perceber, os dias com maior consumo médio são as segundas-feiras e os dias úteis, enquanto os sábados, domingos e feriados apresentam menor consumo. Neste último grupo, os domingos são os dias com menor consumo médio, apresentando ainda menos que os feriados. Estes resultados corroboram a figura anterior, onde se tinham verificado padrões semelhantes.

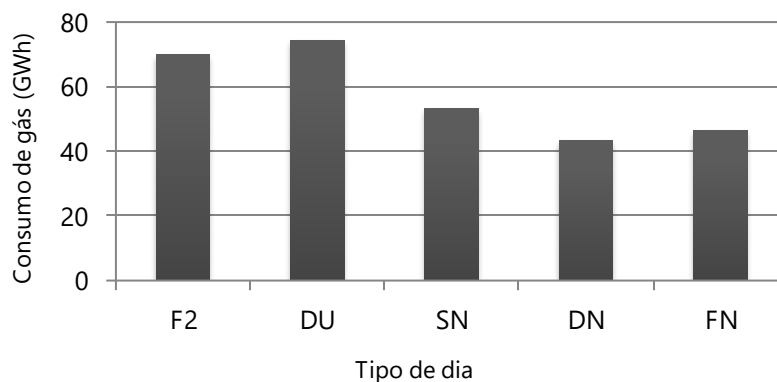


Figura 4.37 – Consumo médio por tipo de dia.

Em relação aos coeficientes de correlação, os valores para cada tipo de dia, apresentados na tabela 4.15, são muito altos, o que evidencia a existência de uma relação forte entre o tipo de dia e o consumo de gás. Os dados por cada tipo de dia têm uma variedade grande de temperaturas e permitem detetar esta associação. Pelo contrário, os dados relativos a uma estação do ano, apresentados na tabela 4.16, têm uma pequena dispersão de temperaturas não permitindo a deteção da relação anterior. A título de exemplo, no que toca ao inverno, os valores das temperaturas são muito baixos e os do consumo muito altos, logo a probabilidade de encontrar uma relação é baixa, como se pode confirmar nas próximas figuras. Os dados referentes aos feriados são muito escassos pelo que o respetivo coeficiente de correlação merece alguma reserva.

Tabela 4.15 – Coeficiente de correlação por tipo de dia.

ρ dias úteis	ρ sábados	ρ domingos	ρ feriados
-0.8007	-0.8522	-0.8607	-0.3925

Tabela 4.16 – Coeficiente de correlação por estação do ano.

ρ inverno	ρ primavera	ρ verão	ρ outono
-0.0262	-0.2496	-0.1025	-0.2308

O que se disse anteriormente é bem visível na figura 4.38, correspondendo aos dias úteis no geral e aos dias de inverno, respetivamente. A gama de temperaturas dos dias de inverno é demasiado baixa para detetar a correlação negativa, a qual é bem visível na figura 4.38 a). Na figura 4.38 b) observam-se alguns pontos dispersos, referentes a consumos mais baixos que, provavelmente, se devem à distinção entre os tipos de dia, a qual não foi tomada em conta.

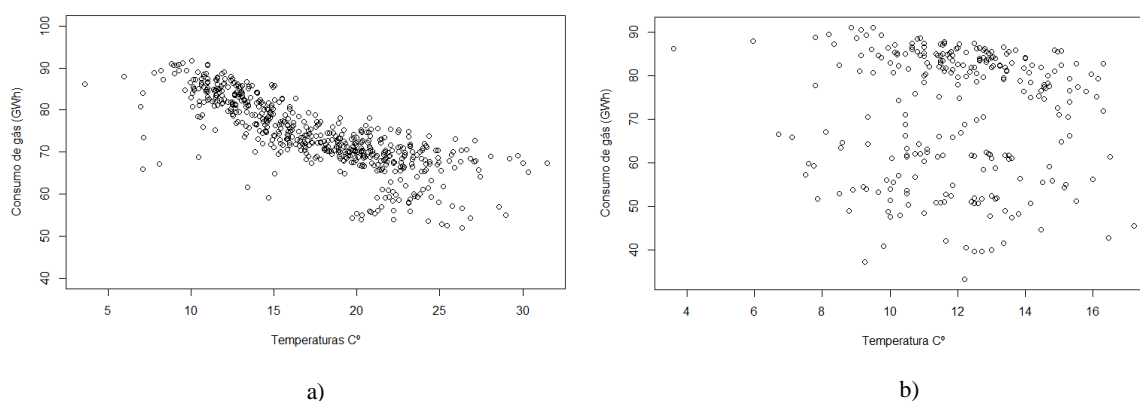


Figura 4.38 – a) Diagrama de dispersão entre o consumo de gás e temperatura para os dias úteis.
b) Diagrama de dispersão entre o consumo de gás e temperatura para o inverno.

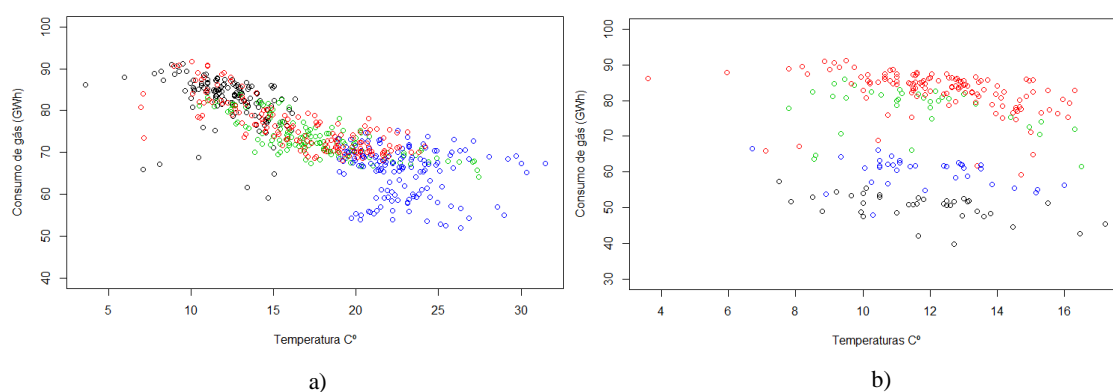


Figura 4.39 – a) Diagrama de dispersão entre o consumo de gás e temperatura para os dias úteis dividida pelas estações do ano. b) Diagrama de dispersão entre o consumo de gás e temperatura para o inverno dividida pelo tipo de dia.

Na figura 4.39 a), apresentam-se os consumos nos dias úteis divididos pelas estações do ano: inverno – preto, primavera – verde, verão – azul e outono – vermelho. O padrão observado é aquele que se esperava, tendo em conta a análise da figura 4.35 (igual a este, mas para todos os dias): nuvem “esférica” no inverno e verão, em patamares diferentes (consumos médios muito diferentes), mas indiciando falta de associação. Já no caso da primavera e do outono, os dados encontram-se mais dispersos na temperatura e é possível observar a correlação consumo/temperatura.

Na figura 4.39 b), observa-se também o padrão esperado, pois os dados relativos aos dias úteis (vermelho) e às segundas-feiras (verde), os dias com mais consumo, situam-se na parte superior do gráfico, enquanto os dados relativos aos sábados (azul) e aos domingos (preto) se dissipam na parte inferior do gráfico. Tal como na análise anterior para cada tipo de dia isolado, verifica-se a mesma relação entre o consumo e a temperatura.

Estes padrões verificam-se também para os restantes tipos de dias e para as restantes estações, não deixando dúvidas sobre o comportamento dos dados.

Assim, confirma-se que existe uma relação suficientemente forte entre o consumo de gás e a temperatura observada e pode-se afirmar que existe uma estrutura de dependência entre as ditas variáveis. Concluimos também ser o efeito de calendário bastante significativo. Podemos dizer então que é adequado experimentar a correção da temperatura ao consumo e aplicar o efeito calendário, tal como foi feito no modelo de previsão para o consumo de electricidade.

b. Adaptação do modelo

Neste subcapítulo vai-se estudar a adaptação do modelo aos dados do consumo de gás de mercado convencional de baixa pressão, referentes ao período de 2012 a 2015, fazendo previsões entre 2013 e 2015. O modelo a ser adaptado é o modelo descrito no capítulo dois, utilizado pela empresa para os dados de consumo de electricidade, com o mesmo processo de correção de temperatura e efeito calendário.

Como para este caso não existe nenhum ponto de partida, como por exemplo a existência de gradientes, adaptamos as funções do subcapítulo II deste capítulo de modo a descobrir os gradientes para as previsões de consumo de gás. Nesta adaptação também se mantém o objetivo de obter a convergência dos gradientes, isto é, obter os gradientes que explicam o comportamento dos desvios. Visto não existirem gradientes iniciais, optou-se por valores iniciais nulos.

As funções utilizadas são as mesmas: *funmoddif*, *gradientes* e *tratamento*. No entanto, foram feitas algumas adaptações. A função principal, *funmoddif*, apenas tem uma pequena alteração no que toca aos graus do polinómio de ajustamento; para os três gradientes apenas se consideram os graus no intervalo de 2 a 5 (exclui-se o grau 0, pois não se querem utilizar os valores iniciais e o grau 1, pelas mesmas razões). Na função *gradientes* não há qualquer alteração. Na função *tratamento* foram feitas algumas mudanças para adaptá-la a um conjunto de dados de tamanho diferente.

Em relação aos valores de correção do efeito calendário, estes foram estimados de novo para o consumo de gás. No entanto, no caso dos dias que não ocorreram (neste histórico pequeno existem certos tipos de dia que nunca aconteceram), foram copiados os valores utilizados para as correções de consumo de eletricidade.

Para os primeiros testes, obtiveram-se os seguintes resultados:

Tabela 4.17 – RMSE de previsão para cada teste, diferenciado o conjunto de temperaturas normais.

	RMSE
Velhas Temperaturas Normais	0.04985
Novas Temperaturas Normais	0.04792

Tabela 4.18 – Erro anual em percentagem de previsão de 2013 até 2015.

Erro anual em % do valor previsto, obtido com valores históricos até 2015		
2013	2014	2015
1.606	-3.113	-0.4044

Verificamos que os RMSEs são bastante baixos considerando as previsões com as novas temperaturas normais (o erro é mais baixo que na eletricidade), o que é uma indicação de que este modelo com o processo de correção e efeito de calendário é adequado para os dados do consumo de gás. No entanto, os erros percentuais anuais não mostram resultados tão positivos comparativamente aos dados de consumo de eletricidade.

Como referido, para estas previsões são utilizados os dados de 2012 até 2015, pelo que as previsões (2013-2015) são influenciadas pelos próprios valores reais, o que pode justificar os valores baixos dos erros de previsão.

Apesar dos erros baixos, os novos gradientes, apresentados na figura 4.40, não correspondem ao esperado pela empresa, isto é, as curvas apresentaram um comportamento demasiado errático. No caso do gradiente 1, cuja curva final é a roxa, obtiveram-se poucas oscilações e uma curva suave que começa e acaba no mesmo sinal. O gradiente 2 é o único com comportamento “expectável”, denotando uma curva bastante suave, sem oscilações, com valores positivos no verão, e negativos no inverno (extremidades). Por fim o gradiente 3 é o caso mais problemático. Além das grandes oscilações que revela, os valores iniciais são muito altos, o que implica que o fator de correção de temperatura, indicado na fórmula 4.2, seja muito alto chegando a valores extremos que poderão não corresponder à realidade.

Na figura 4.43 apresenta-se o gráfico que representa a correção multiplicativa da temperatura, face ao desvio absoluto de temperatura, para a semana 1. Analisando este gráfico, verificamos que a reta referente aos desvios de temperatura maiores que 2 tem um declive muito baixo, o que leva a mudanças muito significativas do consumo face a variações pequenas da temperatura. Na tabela 4.19 apresentam-se os últimos quatro valores de correção (para desvios de temperatura de 5.7°C, 5.8°C,

5.9°C e 6°C), e que estão muito próximos de 20%, o que significa que o consumo corrigido é cerca de 20% do consumo que se verificou. Este tipo de correção não pode acontecer pois o efeito de temperatura no consumo não é tão significativo ao ponto de se observar estes níveis de correção.

Fazendo uma análise às eventuais causas destes valores nos três gradientes, chegou-se à conclusão que este comportamento se deve ao facto de existirem poucos (em alguns casos, nenhuns) pontos para calcular os gradientes. Como já descrito, no caso de não existirem pontos ou quando existir apenas um, considera-se o gradiente nulo, o que irá obviamente influenciar negativamente o comportamento da curva. Em certos casos, verificou-se a existência de apenas dois pontos muito afastados em ordenada, o que motivou um gradiente de ajustamento muito alto em valor absoluto.

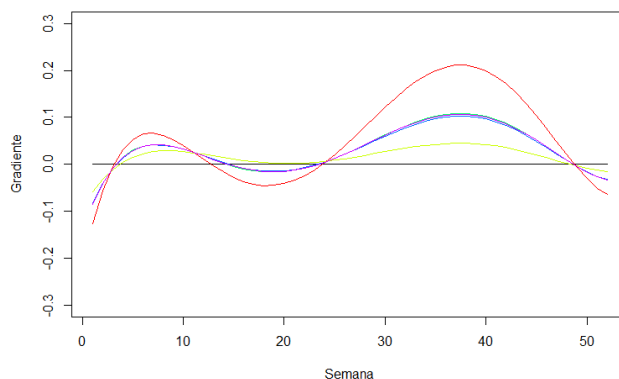


Figura 4.40 – Evolução e convergência do gradiente 1.

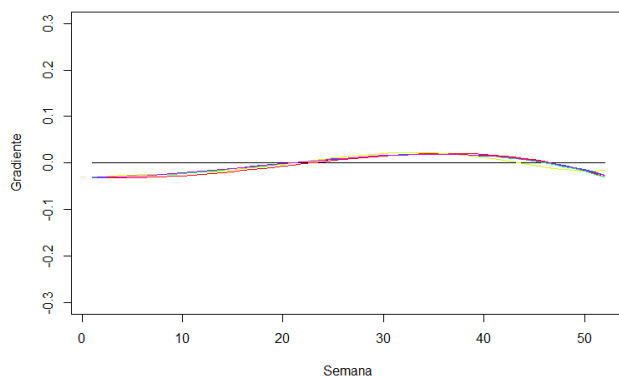


Figura 4.41 – Evolução e convergência do gradiente 2.

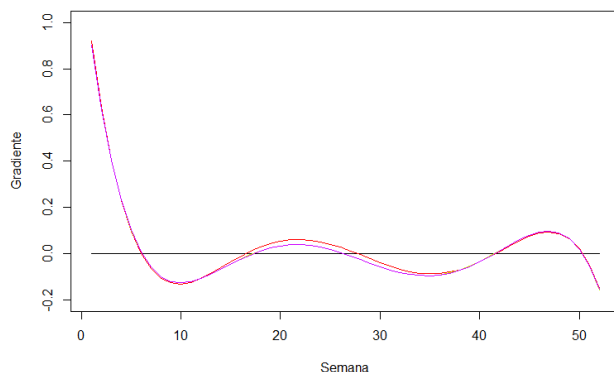


Figura 4.42 – Evolução e convergência do gradiente 3.

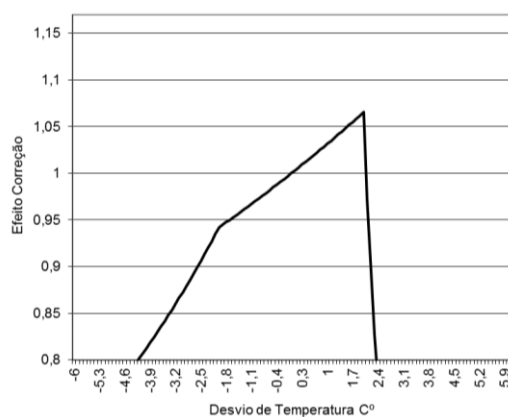


Figura 4.43 – Efeito fator de correção da semana 1.

Tabela 4.19 – Valores de correção direta ao consumo para a semana 1 no caso de um desvio de temperatura maior que 2.

Valores de correção para a semana 1 caso $\phi_{ikj} \geq 2$			
0,233873137	0,229041469	0,224405398	0,219953282

Tendo em conta os resultados anteriores e analisando o motivo do comportamento irregular das curvas, decidiu-se optar por uma nova perspetiva. Como a razão principal para tal resultado consistia na pequena quantidade de pontos, para além dos dados da semana correspondente ao gradiente, decidiu-se seleccionar também os dados da semana anterior e da seguinte, ficando assim com uma maior quantidade de pontos.

Com este novo conjunto de dados, realizaram-se de novo os testes:

Tabela 4.20 – RMSE de previsão para cada teste diferenciado o conjunto de temperaturas normais.

	RMSE
Velhas Temperaturas Normais	0.05049
Novas Temperaturas Normais	0.04751

Tabela 4.21 – Erro anual em percentagem de previsão de 2013 até 2015.

Erro anual em % do valor previsto valores históricos até 2015		
2013	2014	2015
1.576	-2.403	-1.16

Comparando o RMSE, verifica-se uma melhoria com a nova abordagem, suportando a tese de que as temperaturas normais necessitam de uma atualização a fim de trazer benefícios. Os valores de RMSE continuam muito baixos, mesmo que se registre uma ligeiríssima subida com as temperaturas normais “velhas”. Em relação aos erros percentuais, apenas nos anos 2013 e 2014 registamos erros menores, subindo o registado no ano de 2015.

Analisando as seguintes figuras, correspondentes à convergência das curvas dos três gradientes, verificam-se grandes melhorias. Ainda assim, continuam a existir alguns valores altos indesejados, principalmente no gradiente 1.

Estudando a curva final para o gradiente 1, apesar de esta começar e acabar no mesmo sinal (positivo), os valores das pontas diferem bastante, o que leva a que o fator de correção tenha valores muito diferentes de uma semana para a outra (fim de dezembro e início de janeiro), descontinuidade

essa que não é desejável. Relativamente ao gradiente 2 mantém-se o resultado obtido pela abordagem inicial, sem oscilações, revelando um comportamento aceitável. Por fim, o gradiente 3 apresenta bastantes melhorias em relação ao que se tinha obtido na abordagem anterior, com um comportamento mais satisfatório. Porém, começa com um sinal negativo e acaba com um sinal positivo, o que irá influenciar o fator de correção, gerando valores de correção contrários para semanas consecutivas, o que constitui uma mudança brusca indesejável.

Na figura 4.47 observa-se o fator de correção da semana cinquenta e dois, onde a reta para os desvios de temperatura menores que -2 apresenta um declive enorme em valor absoluto quando comparado com os declives das outras retas de regressão. Isso leva a que os desvios percentuais de consumo para certos desvios absolutos de temperatura estejam muito próximos de -1 , ora superiores ora inferiores, e consequentemente os fatores de correção sejam extremamente elevados e instáveis. Na tabela 4.22 apresentam-se os valores do fator de correção para os desvios de temperatura de -6°C , $-5,9^{\circ}\text{C}$, $-5,8^{\circ}\text{C}$ e $-5,7^{\circ}\text{C}$. Os valores apresentados são instáveis e extremos sendo muito altos ou muito baixos, chegando mesmo a atingir valores impossíveis, ou seja, valores negativos.

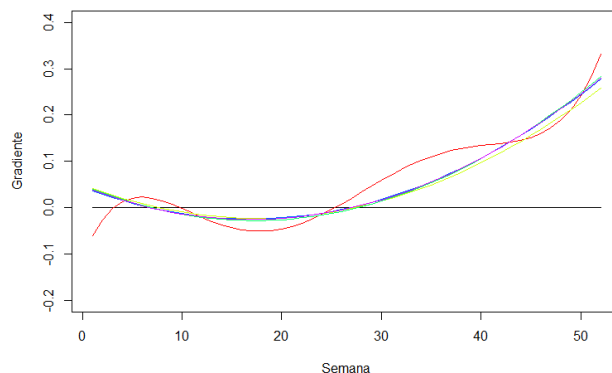


Figura 4.44 – Evolução e convergência do gradiente 1.

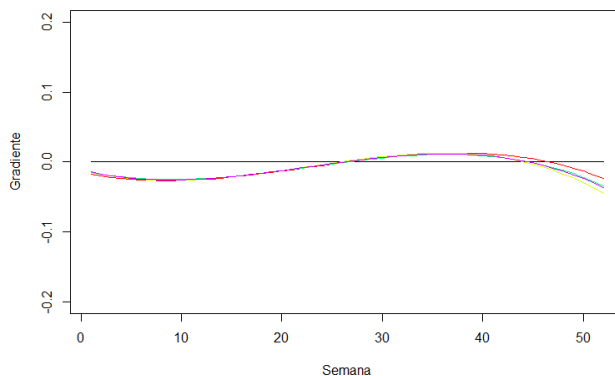


Figura 4.45 – Evolução e convergência do gradiente 2.

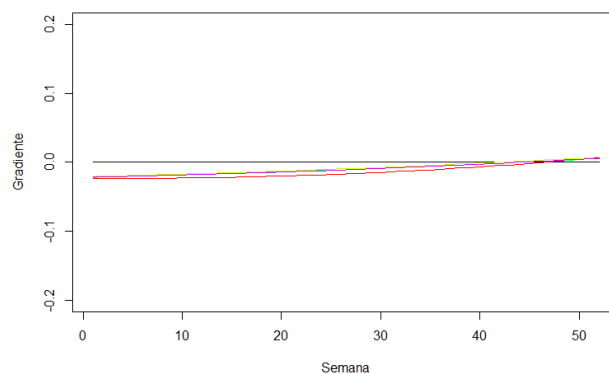


Figura 4.46 – Evolução e convergência do gradiente 3.

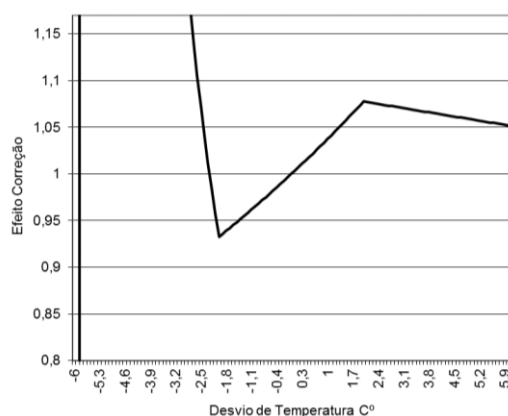


Figura 4.47 – Efeito fator de correção da semana 52.

Tabela 4.22 – Valores de correção direta ao consumo para a semana 52 no caso de um desvio de temperatura menor que -2.

Valores de correção para a semana 52 caso $\phi_{ikj} < -2$			
-22,98639206	-64,05493351	81,42792001	24,8922167

Apesar dos erros baixos ainda existem algumas falhas no comportamento das curvas dos gradientes, que se devem principalmente ao facto de existirem poucos pontos. Por ser apenas possível ter acesso a quatro anos de dados, não se consegue reunir muita informação sobre o comportamento do consumo do gás natural, quando comparado com os treze anos de consumo de eletricidade. Além desta pequena quantidade de pontos, o processo envolvente no modelo poderá não ser o mais correto, pois o comportamento dos dados do consumo de gás é diferente e poder-se-á pensar em rever certas fases do processo de correção de temperatura.

Independentemente destas falhas os erros são baixos, quase tão baixos como os de previsão de consumo de eletricidade o que indica que o modelo é ajustável à situação, sem, no entanto, deixar de pensar que certas fases do mesmo deverão ser revistas e ajustadas.

Capítulo 5 - Conclusão

O mercado da energia é um mercado que necessita de dados com precisão sustentada e que funciona com informações de natureza muito diversa, em regime diário. Com informações no campo do consumo da eletricidade e da respetiva produção, quer como energias não renováveis quer como renováveis, estas devem ser analisadas e estudadas com um grande cuidado.

A REN – Redes Energéticas Nacionais, empresa responsável pela operação do Mercado de Serviços de Sistema, lida com milhões de dados sobre estas energias que vão desde a energia produzida nas inúmeras centrais do país até dados internacionais, de modo a proporcionar a compra e a venda de energia. Aproveitando os métodos de produção de energia espalhados pelo país e evitando custos elevados com a compra de energia, a empresa tem por objetivo atingir o equilíbrio entre o nível de consumo e o nível de produção. Para tal, possui um modelo de previsão, onde se tem como restrição a minimização dos erros de previsão.

O modelo de previsão, descrito anteriormente, engloba várias fases de processamento e inclui os fatores temperatura e tipo de dia. Este projeto surgiu com a intenção de atualizar e redefinir certas fases envolvendo estes fatores fundamentais. Com a evolução das temperaturas (aquecimento global), deixa de fazer sentido utilizar um histórico abrangente (desde 1977) e estudou-se o desenvolvimento do comportamento da relação entre o consumo de eletricidade e as temperaturas face à redução da abrangência do histórico de temperaturas. Neste trabalho a apresentação de resultados consiste no RMSE de previsão, uma medida de desempenho largamente utilizada, e nos erros anuais em percentagem. Foi investigada a hipótese de incluir um novo fator, a luminosidade, através de estudos de correlação. Por fim, com o aumento da importância do gás natural e do crescimento deste no mercado, a empresa decidiu estudar a possível adaptação ao modelo dos dados de consumo de gás.

Ao longo das fases do projeto provou-se que era necessária uma atualização das temperaturas normais e do fator corretivo do consumo face aos desvios da temperatura relativamente ao valor normal, isto se compararmos com os valores que a empresa utilizava no modelo antes da atualização. Estas melhorias verificaram-se pela aplicação das funções para redefinir os gradientes no processo de correção de temperatura, pela obtenção de valores de correção credíveis, por melhores erros de previsão, sejam RMSE ou anuais em percentagem. Assim, considera-se que o objetivo desta primeira fase foi atingido.

Em relação à luminosidade mostrou-se a existência de correlação entre o consumo de eletricidade e a medida de luminosidade, após diversas análises para a descoberta da melhor medida para esta quantidade. Infelizmente o processo de inclusão desta variável no modelo é muito longo, envolvendo a alteração e desenvolvimento de ficheiros ASCII e executáveis, processo que não foi possível executar neste projeto, mas que fará sentido executar no futuro.

Por fim, relativamente ao gás, comprovou-se que existe correlação entre a temperatura e o consumo de gás natural e ainda que o tipo de dia tem alguma influência no consumo. O comportamento da série, com a presença de sazonalidade, indicou que o modelo Holt-Winters era também adequado, pelo que se avançou para testes com este modelo. Nesta fase do projeto também se obtiveram erros baixos (RMSE e erros anuais percentuais) e verificaram-se melhorias através das novas temperaturas normais. No entanto, tal como explicado anteriormente, o comportamento dos gradientes resultou em valores do fator de correção muito instáveis e desadequados, incluindo valores impossíveis, pelo que existe alguma fase no processo de previsão que ainda não é adequada a estes

dados. Comparando os resultados com os de consumo de eletricidade, verifica-se que em ambos os casos os gradientes apresentam gamas de valores aproximadas. Como trabalho futuro poder-se-ão verificar os valores fixos, como o parâmetro λ ou os valores g_{inf} e g_{sup} , ou em vez de se utilizarem três gradientes, utilizar-se apenas um. Uma vez que a principal causa dos resultados menos satisfatórios é a pequena quantidade de pontos, com apenas um gradiente poder-se-á ter um comportamento mais aceitável e consequentemente fatores de correção com valores aceitáveis.

O presente projeto desenvolveu uma metodologia que permitiu melhorar os resultados no modelo de previsão de consumo de eletricidade existente e confirmou igualmente a necessidade de uma atualização das temperaturas normais devido às mudanças climáticas. Adicionalmente, estudou-se a inclusão de uma nova variável de forma a inovar o processo de previsão, no entanto não foi possível concluir este estudo. Por fim, apesar das falhas nas previsões do consumo de gás, os erros obtidos são bastante baixos, pelo que após algumas revisões processuais a averiguar, deverá ser possível a obtenção de um modelo e processo envolvente completos e que seja completamente satisfatório para a empresa.

Referências

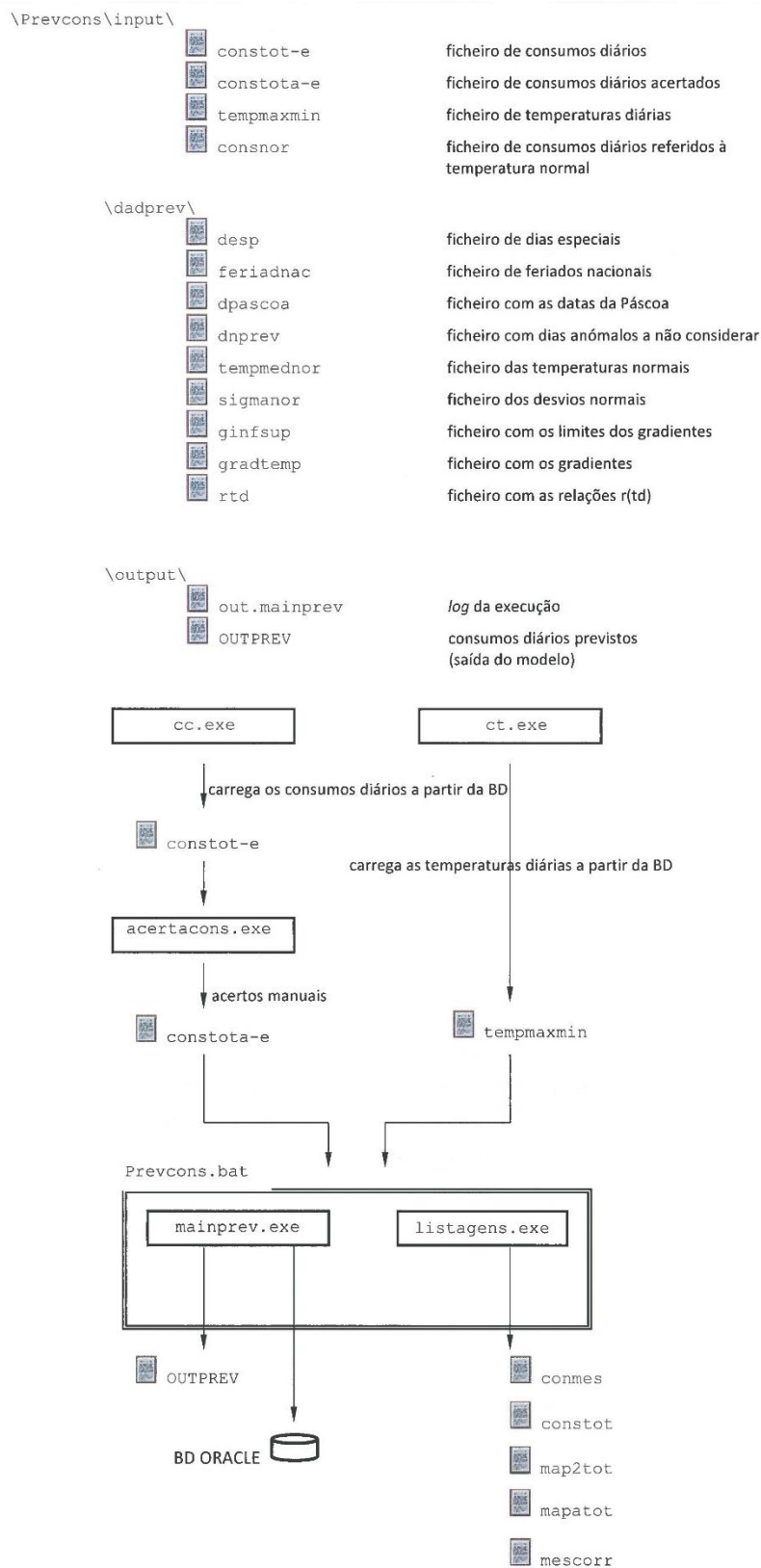
- Energias Renováveis, consultado em junho 2016, em <http://www.ageneal.pt/content01.asp?BtreeID=00/01&treeID=00/01&auxID=&newsID=8&offset=#content>
- Energias Não Renováveis, consultado em junho 2016, em <http://www.ageneal.pt/content01.asp?BtreeID=00/01&treeID=00/01&auxID=&newsID=7&offset=#content>
- Energias Renováveis, consultado em junho 2016, em <http://www.apren.pt/pt/energias-renovaveis/o-que-sao/>
- Gás Natural, consultado em junho 2016, em <http://naturalgas.org/overview/background/>
- REN – Redes Energéticas Nacionais, consultado em Junho 2016, em <http://www.ren.pt/>
- GALP Energia, consultado em junho 2016, em <http://www.galpenenergia.com>
- REN – Informação de Mercado, *PREVCONS, manual de apoio ao utilizador*, 2014
- Bachir, Nádia (2015): “Fatores explicativos na oferta de serviços farmacêuticos em Portugal” – Faculdade de Ciências da Universidade de Lisboa.
- Gomes, João, *Apontamentos Teóricos – Slides Regressão Linear*, 2011.
- Fernandes, Joana T. A. (2012): “Análise de Séries Temporais no Domínio da Frequência. Importância do Periodograma neste contexto” – Faculdade de Ciências da Universidade de Lisboa.
- Machado, Maura A. L. (2012): “Modelos de previsão aplicados à optimização da gestão das actividades de um Call Center” – Faculdade de Ciências da Universidade de Lisboa.
- Alpuim, Teresa, *Detecção de periodicidade em séries temporais: o Periodograma*, 1995
- Rodrigues, A. J., *Processos de Previsão e Decisão – Notas Apoio à Disciplina*, 2015
- Crato, N., *Alisamento Exponencial (EWMA) e Holt-Winters*, ISEG.
- Dias, Rosália (2009): “Implementação de modelos de Previsão da Procura em combustíveis petrolíferos: O caso da Companhia Logística de Combustíveis, S.A.” – Instituto Superior Técnico de Lisboa.
- Lavado, Ana Luísa C. (2009): “Os Actuais Desafios da energia. Implementação e Utilização das Energias Renováveis.” – Faculdade de Ciências da Universidade de Lisboa.
- Botelho, Diogo S. B. S. (2012): “Estratégia de optimização do sistema ibérico de produção de energia eléctrica com grande penetração de energias renováveis.” – Faculdade de Ciências da Universidade de Lisboa.
- Coelho, Rui M. G. A. (2006): “Métodos de *Clustering* para previsão do perfil de carga eléctrica.” – Faculdade de Ciências da Universidade de Lisboa.
- Alpuim, Teresa, *Séries Temporais*, 1997

Anexos

Semanas fixas definidas pela empresa para o modelo de previsão semanal, referido no capítulo 2, subcapítulo III:

Semanas Fixas		
1	01/01	a 07/01
2	08/01	a 14/01
3	15/01	a 21/01
4	22/01	a 28/01
5	29/01	a 04/02
6	05/02	a 11/02
7	12/02	a 18/02
8	19/02	a 25/02
9 (*)	26/02	a 04/03
10	05/03	a 11/03
11	12/03	a 18/03
12	19/03	a 25/03
13	26/03	a 01/04
14	02/04	a 08/04
15	09/04	a 15/04
16	16/04	a 22/04
17	23/04	a 29/04
18	30/04	a 06/05
19	07/05	a 13/05
20	14/05	a 20/05
21	21/05	a 27/05
22	28/05	a 03/06
23	04/06	a 10/06
24	11/06	a 17/06
25	18/06	a 24/06
26	25/06	a 01/07
27	02/07	a 08/07
28	09/07	a 15/07
29	16/07	a 22/07
30	23/07	a 29/07
31	30/07	a 05/08
32	06/08	a 12/08
33	13/08	a 19/08
34	20/08	a 26/08
35	27/08	a 02/09
36	03/09	a 09/09
37	10/09	a 16/09
38	17/09	a 23/09
39	24/09	a 30/09
40	01/10	a 07/10
41	08/10	a 14/10
42	15/10	a 21/10
43	22/10	a 28/10
44	29/10	a 04/11
45	05/11	a 11/11
46	12/11	a 18/11
47	19/11	a 25/11
48	26/11	a 02/12
49	03/12	a 09/12
50	10/12	a 16/12
51	17/12	a 23/12
52	24/12	a 31/12

Ficheiros ASCII e ficheiros executáveis para a execução do modelo e diagrama de processos computacionais referido no capítulo 2, subcapítulo III:



Desvio-padrão anual das temperaturas normais, referido no capítulo 2, subcapítulo III:

Desvio Padrão das Temperaturas Normais ^(*)												°C
DIAS	Janeiro	Fevereiro	Março	Abril	Maió	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
1	2.37	2.02	2.16	2.35	2.52	2.70	2.82	2.78	2.52	2.20	2.07	2.29
2	2.36	2.02	2.16	2.35	2.52	2.71	2.82	2.77	2.51	2.19	2.07	2.30
3	2.31	2.02	2.17	2.36	2.53	2.71	2.82	2.76	2.50	2.18	2.07	2.31
4	2.28	2.02	2.18	2.36	2.54	2.72	2.83	2.76	2.49	2.17	2.07	2.32
5	2.25	2.03	2.18	2.37	2.54	2.72	2.83	2.75	2.47	2.16	2.08	2.33
6	2.23	2.03	2.19	2.37	2.55	2.73	2.83	2.75	2.46	2.15	2.08	2.35
7	2.21	2.03	2.20	2.38	2.55	2.73	2.83	2.74	2.45	2.15	2.09	2.36
8	2.18	2.04	2.20	2.39	2.56	2.74	2.83	2.73	2.44	2.14	2.09	2.37
9	2.16	2.04	2.21	2.39	2.57	2.74	2.83	2.73	2.43	2.13	2.10	2.37
10	2.15	2.05	2.21	2.40	2.57	2.75	2.83	2.72	2.42	2.13	2.10	2.38
11	2.13	2.05	2.22	2.40	2.58	2.75	2.83	2.71	2.41	2.12	2.11	2.39
12	2.11	2.06	2.23	2.41	2.58	2.76	2.83	2.71	2.40	2.11	2.11	2.40
13	2.10	2.06	2.23	2.41	2.59	2.76	2.83	2.70	2.39	2.11	2.12	2.41
14	2.09	2.07	2.24	2.42	2.60	2.77	2.83	2.69	2.37	2.10	2.13	2.42
15	2.08	2.07	2.25	2.43	2.60	2.77	2.83	2.68	2.36	2.10	2.13	2.42
16	2.07	2.08	2.25	2.43	2.61	2.78	2.83	2.67	2.35	2.09	2.14	2.43
17	2.06	2.08	2.26	2.44	2.61	2.78	2.82	2.66	2.34	2.09	2.15	2.43
18	2.05	2.09	2.26	2.44	2.62	2.78	2.82	2.66	2.33	2.08	2.16	2.44
19	2.04	2.10	2.27	2.45	2.63	2.79	2.82	2.65	2.32	2.08	2.17	2.44
20	2.03	2.10	2.28	2.45	2.63	2.79	2.82	2.64	2.31	2.07	2.18	2.44
21	2.03	2.11	2.28	2.46	2.64	2.79	2.82	2.63	2.30	2.07	2.19	2.44
22	2.02	2.11	2.29	2.47	2.64	2.80	2.81	2.62	2.29	2.07	2.20	2.44
23	2.02	2.12	2.29	2.47	2.65	2.80	2.81	2.61	2.28	2.07	2.21	2.44
24	2.02	2.13	2.30	2.48	2.66	2.80	2.81	2.60	2.27	2.07	2.22	2.44
25	2.02	2.13	2.31	2.48	2.66	2.81	2.80	2.59	2.26	2.06	2.23	2.44
26	2.01	2.14	2.31	2.49	2.67	2.81	2.80	2.58	2.25	2.06	2.24	2.43
27	2.01	2.14	2.32	2.49	2.67	2.81	2.80	2.57	2.24	2.06	2.25	2.42
28	2.01	2.15	2.32	2.50	2.68	2.82	2.79	2.56	2.23	2.06	2.26	2.42
29	2.01	2.16	2.33	2.51	2.69	2.82	2.79	2.55	2.22	2.06	2.27	2.40
30	2.01		2.33	2.51	2.69	2.82	2.78	2.54	2.21	2.06	2.28	2.39
31	2.01		2.34		2.70		2.78	2.53		2.06		2.38

(-) ajuste polinomial à curva de desvios padrão das temperaturas médias diárias registadas em Lisboa/Portela, desde 1977

Função principal *funmoddif*, referida no capítulo 4, subcapítulo I:

```

fic1='out'
fic2='consv'
fic3='tdesv'
options("scipen"=100, "digits"=4)

funmoddif=function(fic1,fic2,fic3){

  fic1=paste0('D:/Prevcons/temp/',fic1,'.csv')
  fic2=paste0('D:/Prevcons/',fic2,'.csv')
  fic3=paste0('D:/Prevcons/',fic3,'.csv')
  system("D:/Prevcons/Prevcons.bat",wait=T)

  dados=tratamento(fic1,fic2,fic3)

  ograd=read.table("D:/Prevcons/ngt-e")

  temp=dados[-(1:365),]
  dt=sqrt(sum(temp$desv^2)/length(temp$desv))
  grad=gradientes(dados)

  it=1
  o1=F
  o2=F
  o3=F

  newgrad1=data.frame(rep(0,52))

```

```

newgrad2=data.frame(rep(0,52))
newgrad3=data.frame(rep(0,52))
final=as.data.frame(matrix(0,52,3))
erro=data.frame(c(0))
erro[1,]=dt
gr=rep(0,3)

while(1){

  dc=Inf

  if(o1==FALSE){
    plot(grad[,1],main="grad1",xlab="",ylab="")
    o1=as.logical(readline())
  }
  if(o2==FALSE){
    plot(grad[,2],main="grad2",xlab="",ylab="")
    o2=as.logical(readline())
  }
  if(o3==FALSE){
    plot(grad[,3],main="grad3",xlab="",ylab="")
    o3=as.logical(readline())
  }

  for(i in c(0,2:5)){

    for(j in 2:5){

      for(k in 2:5){

        x=1:52

        if(o1==T){
          new1=final[,1]
        }
        else{
          if(i==0){
            new1=ograd[,1]
          }
          else{
            fit1=lm(grad[,1]~poly(x,i))
            new1=round(fit1$fitted.values,6)
          }
        }

        if(o2==T){
          new2=final[,2]
        }
        else{
          fit2=lm(grad[,2]~poly(x,j))
          new2=round(fit2$fitted.values,6)
        }

        if(o3==T){
          new3=final[,3]
        }
        else{
          fit3=lm(grad[,3]~poly(x,k))
          new3=round(fit3$fitted.values,6)
        }

        newgrad=cbind(new1,new2,new3)

        file.remove("D:/Prevcons/dadprev/ngt-e")
        write.table(newgrad, file="D:/Prevcons/dadprev/ngt -
e",row.names=F,col.names=F, sep=" ")
      }
    }
  }
}

```

```

file.remove("D:/Prevcons/temp/out.csv")
system("D:/Prevcons/Prevcons.bat",wait=T)

dados=tratamento(fic1,fic2,fic3)

temp=dados[-(1:365),]
d=sqrt(sum(temp$desv^2)/length(temp$desv))

if(d<dc){
  dc=d
  finalgrad=newgrad
  data=dados
  grau=c(i,j,k)
}
}
}

newgrad1[,it]=finalgrad[,1]
newgrad2[,it]=finalgrad[,2]
newgrad3[,it]=finalgrad[,3]
erro[it+1,]=dc

if(o1==FALSE){
  plot(ograd[,1],type="l",ylim=c(-0.05,0.02),main=paste0("Evolução gradiente
",1),xlab="",ylab="")
  color=rainbow(dim(newgrad1)[2])
  for(j in 1:dim(newgrad1)[2]){
    lines(newgrad1[,j],col=color[j]) #mudar a cor a cada iteração
  }

  o1=as.logical(readline())
  if(o1==TRUE){
    final[,1]=newgrad1[,dim(newgrad1)[2]]
    gr[1]=grau[1]
  }
}

if(o2==FALSE){
  plot(ograd[,2],type="l",ylim=c(-0.1,0.06),main=paste0("Evolução gradiente
",2),xlab="",ylab="")
  color=rainbow(dim(newgrad2)[2])
  for(j in 1:dim(newgrad2)[2]){
    lines(newgrad2[,j],col=color[j]) #mudar a cor a cada iteração
  }

  o2=as.logical(readline())
  if(o2==TRUE){
    final[,2]=newgrad2[,dim(newgrad2)[2]]
    gr[2]=grau[2]
  }
}

if(o3==FALSE){
  plot(ograd[,3],type="l",ylim=c(-0.1,0.1),main=paste0("Evolução gradiente
",3),xlab="",ylab="")
  color=rainbow(dim(newgrad3)[2])
  for(j in 1:dim(newgrad3)[2]){
    lines(newgrad3[,j],col=color[j]) #mudar a cor a cada iteração
  }

  o3=as.logical(readline())
  if(o3==TRUE){
    final[,3]=newgrad3[,dim(newgrad3)[2]]

```

```

    gr[3]=grau[3]
  }
}

if(o1==TRUE && o2==TRUE && o3==TRUE){
  break
}
else{
  file.remove("D:/Prevcons/dadprev/ngt-e")
  write.table(finalgrad,file="D:/Prevcons/dadprev/ngt-e",row.names=F,
              col.names=F,sep=" ")

  file.remove("D:/Prevcons/temp/out.csv")
  system("D:/Prevcons/Prevcons.bat",wait=T)

  dados=tratamento(fic1,fic2,fic3)
  temp=dados[-(1:365),]
  dt=sqrt(sum(temp$desv^2)/length(temp$desv))

  grad=gradientes(dados)
  it=it+1

  if(o1==T){
    grad[,1]=final[,1]
  }
  if(o2==T){
    grad[,2]=final[,2]
  }
  if(o3==T){
    grad[,3]=final[,3]
  }
}
}

file.remove("D:/Prevcons/dadprev/ngt-e")
write.table(final,file="D:/Prevcons/dadprev/ngt-e",row.names=F,
            col.names=F,sep=" ")

file.remove("D:/Prevcons/temp/out.csv")
system("D:/Prevcons/Prevcons.bat",wait=T)

dados=tratamento(fic1,fic2,fic3)
temp=dados[-(1:365),]
dt=sqrt(sum(temp$desv^2)/length(temp$desv))

return(list(dados,final,dt,gr))
}

```

Função principal *gradientes*, referida no capítulo 4, subcapítulo I:

```

gradientes=function(dados){

  dados=dados[-(1:365),]
  grad=data.frame(rep(0,52))

  for(i in 1:52){
    sem=dados[which(dados$coef==1),]
    sem=sem[which(sem$semana==i),]

    semt2=sem[which(sem$temp<(-2)),]
    if(dim(semt2)[1]==0 || dim(semt2)[1]==1){
      grad[i,1]=0 #PROBLEMA NA
    }
    else{
      fitt2=lm(semt2$desv~semt2$temp, data=semt2)
    }
  }
}

```

```

    grad[i,1]=fitt2$coefficients[2]
  }

  semt02=sem[which(sem$temp<2 & sem$temp>(-2)),]
  if(dim(semt02)[1]==0 || dim(semt02)[1]==1){
    grad[i,2]=0 #PROBLEMA NA
  }
  else{
    fitt02=lm(semt02$desv~semt02$temp)
    grad[i,2]=fitt02$coefficients[2]
  }

  semtm2=sem[which(sem$temp>2),]
  if(dim(semtm2)[1]==0 || dim(semtm2)[1]==1){
    grad[i,3]=0 #PROBLEMA NA
  }
  else{
    fittm2=lm(semtm2$desv~semtm2$temp)
    grad[i,3]=fittm2$coefficients[2]
  }
}
return(grad)
}

```

Função principal *tratamento*, referida no capítulo 4, subcapítulo I:

```

tratamento=function(fic1,fic2,fic3){

  w1=read.csv2(file=fic1,header=F)
  w2=read.csv2(file=fic2,header=F)
  t=read.csv2(file=fic3,header=F)
  w1=w1[, -dim(w1)[2]]
  if(w1$V1[dim(w1)[1]]=="2016-12-31"){
    u=dim(w1)[1]-365
    v=dim(w1)[1]
    w1=w1[-(u:v),]
  }
  w2=as.data.frame(w2[, -c(1,dim(w2)[2])])

  dados=cbind(w1,w2,t)
  names(dados)=c("data","semana","prevf","prev","coef","cons","temp")
  dados$data=as.character(dados$data)
  dados$prevf=as.numeric(paste(dados$prevf))
  dados$prev=as.numeric(paste(dados$prev))
  dados$coef=as.numeric(paste(dados$coef))
  dados$cons=as.numeric(paste(dados$cons))
  dados$temp=as.numeric(paste(dados$temp))
  dados$desv=(dados$cons-dados$prevf)/dados$cons

  return(dados)
}

```